TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

TOPI JÄRVENSIVU

DEFINITION AND MEASUREMENT OF SUPPLIER QUALITY IN INDUSTRIAL SERVICES

Master of Science thesis

Examiners: Prof. Jussi Heikkilä and Senior Research Fellow Aki Jääskeläinen
Examiners and topic approved by the Faculty Council of the Faculty of Business and Technology
on 9th September 2016

ABSTRACT

TOPI JÄRVENSIVU: DEFINITION AND MEASUREMENT OF SUPPLIER QUALITY IN INDUSTRIAL SERVICES
Tampere University of Technology
Master of Science Thesis, 109 pages, 11 Appendix pages
May 2017
Master's Degree Programme in Industrial Engineering and Management
Major: Production control and logistics
Examiners: Professor Jussi Heikkilä and Senior Research Fellow Aki Jääskeläinen

Keywords: service quality, purchasing, industrial services, service profit chain, supplier service quality measurement, survey

Services have been traditionally separated from goods in terms of intangibility, heterogeneity, inseparability and perishability. Even though the line between goods and services is blurred, as many services include a physical component and vice versa, services pose some differences to the purchasing process compared to tangible goods. With services, the importance of the specification phase is highlighted due to the fact that services are difficult to define unambiguously. This has also implications on the performance evaluation of the service supplier: the difficulty of defining a service makes it challenging to measure service quality.

The objective of this research was to develop an approach for supplier quality measurement in industrial services from the viewpoint of the buyer. However, the intention was that the measurement system would ultimately be used jointly by the buyer and the supplier. The research in this thesis was conducted as a case study utilizing a mixed-method research approach. The empirical part of the research consists of two parts: the definition and measurement of supplier quality in industrial services. In defining supplier service quality, literature review as well as interviews with the case company and the service supplier representatives were used. As a result, a framework for supplier service quality was developed. The framework offers a comprehensive view on supplier service quality, combining the viewpoints of service profit chain, relationship quality, and process and outcome quality of the service. Based on the framework, supplier service quality consists of four dimensions: supplier capability, supplier-customer relationship, and process and outcome quality.

The measurement of supplier service quality was constructed as a survey based on the developed framework, and it was used to measure the quality of cleaning service. This thesis describes the undergone process from the development of the measurement items to the actual data gathering and analyzing the results. Statistical analysis was used to examine the survey data. This thesis also developed a model of supplier service quality, aiming to examine the links between the four service quality dimensions. The results provide support for the use of process and outcome quality as dimensions of service quality. Even though the model could not be examined in full, the results suggest that process quality has an effect on outcome quality. Overall, the measurement process reported in this thesis offers useful insights for the future use and development of the supplier service quality measurement.

TIIVISTELMÄ

TOPI JÄRVENSIVU: TOIMITTAJAN LAADUN MÄÄRITTELY JA MITTAAMINEN TUOTANNOLLISISSA PALVELUISSA
Tampereen teknillinen yliopisto
Diplomityö, 109 sivua, 11 liitesivua
Toukokuu 2017
Tuotantotalouden diplomi-insinöörin tutkinto-ohjelma
Pääaine: Tuotannonohjaus ja logistiikka
Tarkastajat: Professori Jussi Heikkilä ja Yliopistotutkija Aki Jääskeläinen

Avainsanat: palvelun laatu, hankinta, tuotannolliset palvelut, palvelun tuottoketju, toimittajan palvelun laadun mittaaminen, kysely

Palvelut on perinteisesti erotettu fyysisistä tuotteista käyttäen neljää ominaisuutta: aineettomuus, heterogeenisuus, tuotannon ja kulutuksen samanaikaisuus ja katoavuus. Rajanveto palveluiden ja tuotteiden välille on vaikeaa, kun monet palvelut sisältävät myös fyysisiä elementtejä, ja päinvastoin. Palvelut kuitenkin asettavat erilaisia vaatimuksia hankintaprosessille kuin fyysiset tuotteet. Määrittelyvaiheen tärkeys korostuu erityisesti palveluissa, koska palveluita on vaikea määritellä yksiselitteisesti. Tämä vaikuttaa myös palvelutoimittajan suorituskyvyn arviointiin: palvelun määrittelyn vaikeus tekee palvelun laadun mittaamisen haasteelliseksi.

Tämän työn tavoitteena oli kehittää lähestymistapa toimittajan laadun mittaamiseen tuotannollisissa palveluissa ostajayrityksen näkökulmasta. Kehitettyä mittausjärjestelmää on kuitenkin lopulta tarkoitus käyttää yhteisesti ostajan ja toimittajan kanssa. Tämä tutkimus toteutettiin tapaustutkimuksena käyttäen sekamenetelmätutkimusta. Tutkimuksen empiirinen osuus koostuu kahdesta osasta: toimittajalaadun määrittelystä ja mittaamisesta tuotannollisissa palveluissa. Toimittajan palvelun laadun määrittely pohjautui kirjallisuuskatsaukseen sekä kohdeyrityksen ja toimittajan edustajien haastatteluihin. Määrittelyn tuloksena kehitettiin toimittajan palvelun laadun viitekehys. Viitekehys antaa kokonaisvaltaisen kuvan toimittajan palvelun laadusta, yhdistäen palvelun tuottoketju –ajattelun, yhteistyösuhteen laadun sekä palvelun prosessin ja lopputuloksen laadun. Viitekehyksen mukaan toimittajan palvelun laatu koostuu neljästä ulottuvuudesta: toimittajan kyvykkyydestä, toimittaja-asiakas yhteistyösuhteesta, sekä prosessin ja lopputuloksen laadusta.

Toimittajan palvelun laadun mittaus toteutettiin viitekehykseen pohjautuen kyselynä, ja palveluksi valittiin siivouspalvelu. Tämä diplomityö raportoi läpikäydyn prosessin kyselyväittämien kehityksestä vastausten keräämiseen ja kyselytulosten analysointiin. Kyselydatan analysoinnissa hyödynnettiin tilastollista analyysiä. Työssä kehitettiin myös malli toimittajan palvelun laadulle, jonka tarkoituksena oli tutkia palvelun laadun ulottuvuuksien keskinäisiä suhteita. Tulokset tukevat prosessin ja lopputuloksen laadun käyttämistä palvelun laadun ulottuvuuksina. Vaikka mallia ei voitu tutkia kokonaisuudessaan, tulokset viittaavat siihen, että prosessin laadulla on vaikutusta lopputuloksen laatuun. Kokonaisuudessaan tässä työssä raportoitu mittausprosessi tarjoaa hyödyllisiä näkemyksiä toimittajan palvelun laadun mittaamisen kehittämiseen ja hyödyntämiseen tulevaisuudessa.

PREFACE

This thesis has been the single biggest and most interesting project in my life so far. As such, it has been an invaluable experience in preparing me to finish my studies and start my career. I would like to thank everyone involved in the making of this thesis. Especially, I would like to thank professor Jussi Heikkilä and senior research fellow Aki Jääskeläinen for their support and guidance throughout this project. Also the help of the research staff of the good old Faculty of Industrial Management is acknowledged. A special mention goes to the representatives of Metsä Group for their openness and assistance throughout the project.

Tampere, 19.5.2017

Topi Järvensivu

CONTENTS

LIST OF FIGURES

LIST OF TERMS AND ABBREVIATIONS

CFA	Confirmatory factor analysis
CFI	Comparative fit index
CMV	Common method variance
EFA	Exploratory factor analysis
IHIP	Characteristics of services: intangibility, heterogeneity, inseparability and perishability
KMO	Kaiser-Meyer-Olkin measure of sampling adequacy
KPI	Key performance indicator
MRRS	Mill related support services
OM	Operations Management
PSM	Purchasing and supply management
R^2	Squared multiple correlation coefficient
RMSEA	Root mean square error of approximation
SEM	Structural equation modeling
SERVQUAL	A service quality model developed by Parasuraman et al. (1988)
SLA	Service level agreement
SPC	Service profit chain

1. INTRODUCTION

1.1 Background and motivation

Purchasing is recognized as an important function in organizations today. The importance of purchasing can be seen when looking at the cost structures of companies: The largest part of cost of goods sold is taken by purchased materials and services. Purchased materials and services constitute about 60 to 80 percent of cost of goods sold, and that share is growing. Therefore, it is easy to see the impact of purchasing on the company's profit. (Heikkilä et al. 2013, p. 10.) This also shows how much companies are dependent on their suppliers. The need to manage service procurement is important also because suppliers increasingly provide value-added services to their customers (Sheth 1996, p. 14). Specialization, i.e. companies concentrating on core competencies, has resulted in companies outsourcing tasks that they previously performed in-house. This implies that less added value is created internally by each company. (Axelsson & Wynstra 2002, p. 9, 20.) Therefore, this externally added value needs to be managed effectively, as suppliers have a direct impact on e.g. the cost, quality and profits of the buying company (Krause & Scannell 2002, p. 14).

Services and their characteristics have been widely studied, especially during the 1970s (Cook et al. 1999, p. 321). However, there is still no agreed upon definition of service (Grönroos 2007, p. 51-52). The most popular to way describe services has been the IHIP characteristics, i.e. intangibility, heterogeneity, inseparability and perishability. These have been used to separate services from tangible goods (Zeithaml et al. 1985, p. 33). As much of the literature on services has focused on definitional issues and the aspects that distinguish them from goods, there has been a demand for a change of perspective. On the perceived difference of services compared to goods, Levitt (1972, pp. 41-42) states: "There are no such things as service industries. There are only industries whose service components are greater or less than those of other industries. Everybody is in service."

The literature in purchasing management has traditionally focused on the sourcing of goods. Moreover, the predominant focus in the service literature has been on consumer services (b2c) rather than business-to-business (b2b) context. (Wynstra et al. 2006, p. 475.) Purchasing services has generally been considered to be different and more complex than purchasing goods (van Weele 2014, p. 78; Fitzsimmons et al. 1998, p. 372). Especially, measuring the quality of services is more difficult than in the case of tangible goods.

Abundant research and literature exists concerning service quality and service quality measurement. Most of it focuses on service quality in consumer context (b2c) (Gounaris 2005, p. 421), and though the same measures are often not directly applicable to business-to-business (b2b) environment (Durvasula et al. 1999, p. 146), service quality research in consumer markets can still offer helpful insights. One major difference in b2b and b2c markets is that in b2b services the buyer and the service consumer are generally not the same (Smeltzer & Ogden 2002, p. 55). In business-to-business markets the service is generally acquired by the purchasing function, but the end customer is often some other function or personnel group of the buying company. This has implications on the measurement of service quality, and emphasizes the purchasing function's importance: its objective is to ensure the value-creation of the purchased service for the buying company, without forgetting the needs of the end users of the service.

In the existing service quality models the emphasis has been on measuring service quality from the supplier's perspective rather than from the buyer's. From the buying company's point of view this is not an ideal situation considering the management of the purchased service: the specific needs of the buying company may be overlooked. Furthermore, in the service quality literature the focus is on delivering the best possible service quality to the customer. However, in business-to-business services the customer's needs dictate the appropriate level of service quality, and the purchasing function aims to make sure that the buying company does not pay for extra quality. The underlying motivation for this research is therefore to provide further understanding of supplier service quality measurement from the buyer's perspective in b2b industrial services. Especially, this research aims to find a common approach for service quality measurement between the buyer and supplier. This has not received much attention in the literature.

1.2 Research questions and objectives

The objective of this thesis is twofold. First, this thesis explores the factors from which purchased industrial service quality consists of, and the connection between those factors. Second, this thesis aims to explain how purchased industrial service quality can be measured and what kind of process it requires. The data for this research is gathered mainly from purchased cleaning service. Still, the developed approach should be applicable to industrial services in general.

This research aims to develop an approach for supplier quality measurement between the buyer and the supplier company. This research adopts the perspective of the buying company, but aims to incorporate both the supplier's and the buyer's view on service quality. Usually the issue of quality is addressed from either the supplier's or the buyer's point of view. In addition, the buyer and the supplier may measure the service quality separately and with somewhat different measures, which makes it challenging to discuss about the quality and delivery of the service. It also makes it difficult to develop the service in collaboration. The ultimate goal of the case company is to have one common

measurement system for service quality between the buyer and the supplier. This way both organizations will have the same information concerning the service, and the service delivery process and service quality are more transparent. The research undergone in this thesis acts as a starting point for the common measurement system.

The two main research questions of this thesis can be expressed as follows:

1. *What kind of quality information is relevant in the management of industrial service suppliers?*
2. *How can a survey be used to measure the subjective supplier quality in industrial services?*

From the first research question two sub-questions can be derived. These are:

What are the quality-related information needs of purchasing?

What are the factors of quality and how can they be connected?

From the second research question two sub-questions can be derived. These are:

How can the factors of quality be measured to satisfy purchasing information needs?

How can the results of supplier service quality evaluation be presented and what is the status of supplier quality in cleaning service?

By answering these questions, this thesis aims to introduce an approach for supplier service quality measurement. As a result, this thesis documents the undergone process for developing the approach, enabling the future application to other purchased services.

1.3 Research context and the case company

The thesis is done as a part of ProcuValue (Value Creating Procurement) research project. The objective of the research project is to develop new know-how concerning strategic purchasing. The project also aims to generate measurement data to support purchasing. The ultimate goal of the project is to build the sustainable competitiveness of organizations. Four companies are involved in the ProcuValue project: Metsä Group, Posti, Tieto and Valmet. This thesis examines the purchasing of industrial services and is the last case study in the ProcuValue project. This thesis is especially concerned with measuring supplier service quality in industrial services in a way that enables the development of the service and cooperation with suppliers in the future.

The case company of this research is Metsä Group, who has had a significant influence on the topic and contents of this thesis. Metsä Group is a Finnish forest industry group operating in 30 countries. Metsä Group consists of five different business areas: Metsä

Forest, Wood, Fibre, Board and Tissue. Metsä Group is owned by approximately 116 000 forest owners through its parent company Metsäliitto Cooperative. In 2015, Metsä Group had sales worth of 5.0 billion euros and approximately 9 600 employees. (Metsä Group 2016a.) This research is done in close collaboration with Metsä Fibre, which has four pulp mills in Finland: In Joutseno, Kemi, Rauma and Äänekoski. Metsä Fibre's main products are bleached softwood and birch pulp. Metsä Fibre had sales worth of 1.4 billion euros and about 850 employees in 2015. All four pulp mills produced over 2.3 million tons of pulp combined in 2015, for which they used about 12.6 million cubic meters of wood. (Metsä Fibre 2015.) The Rauma pulp mill was used as a basis for the data collection in the first phase of the research, in which the framework for supplier service quality was constructed. In the empirical phase of the research, six of Metsä Group's production units were included in the analysis.

Annual external purchases of Metsä Group are over 2 billion euros. The purchases in Metsä Group are divided into 17 main categories, including Basic chemicals, Pigments and Mill related support services (MRSS). Each of the main categories has its own leader, and Metsä Group's purchasing as a whole is led by Chief procurement officer (CPO). The cleaning service of the production units, from which the data is gathered in this thesis, is a part of the Mill related support services. (Metsä Group 2016b.) The purchased cleaning service is described more specifically in section 4.1.

The need to measure the service quality of suppliers had already been acknowledged in the case company. One study concerning the service quality of a key supplier for the case company has been conducted as a part of an MBA degree (Hyppänen 2015). This thesis further continues and extends that study.

1.4 Research philosophy and process approach

When considering the questions about the research philosophy, approach, and strategy, among other things, Saunders et al. (2009, p. 108) describe the alternative choices using a research onion: the different choices that must be made are presented as separate layers of the onion. The research onion and the respective choices for this research are presented in Figure 1.

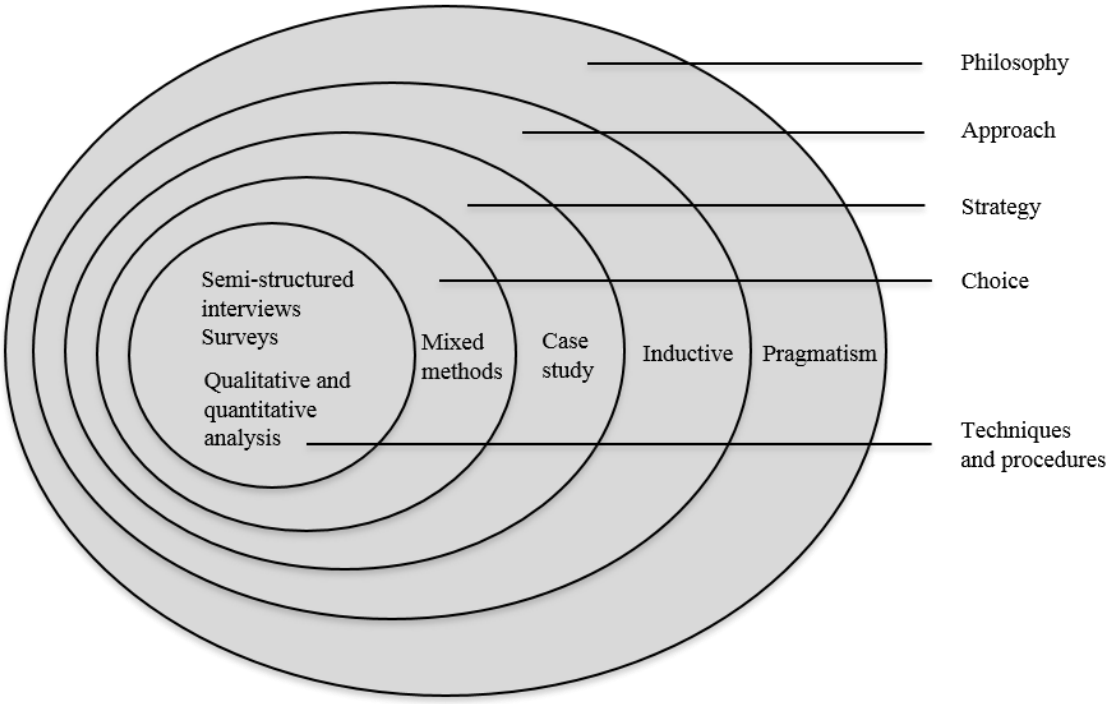


Figure 1. *The research onion and the respective choices of this research (adapted from Saunders et al. 2009, p. 108).*

The research philosophy concerns the way in which the researcher views the world and knowledge (Saunders et al. 2009, p. 108). From the research questions and sub-questions described in Section 1.3, it can be seen that both qualitative and quantitative methods are likely needed to answer them. This emphasizes the role of the research questions. The most suitable research philosophy in this case is pragmatism, where the research questions are treated as most important and the research methods should be chosen accordingly. Especially, it is possible to use variations in epistemology, ontology and axiology, since the appropriate choice depends on the particular question. (Saunders et al. 2009, p. 109.)

The research approach in this research is mainly inductive: a model for supplier service quality is built based on the analysis of the gathered data (Saunders et al. 2009, p. 124). The literature review and interviews were used to define the elements from which the supplier service quality consists of. Countless models on service quality exist, but none of them fully correspond to the objective of this research, which was to develop a basis for a common service quality measurement between the buyer and the supplier. Therefore, testing an existing theory would not have suited the purpose of this research. Still, the literature offered a useful starting point for the supplier service quality measurement, and the developed framework actually combines different models and views on service quality. In that way this research can also be seen as theory refinement. Moreover, the factors that were discovered in the interviews as affecting the supplier service quality, were all backed up by evidence from the literature. A particular strength of the inductive approach is developing an understanding of how humans interpret their

social world, and the inclusion of that understanding in to the studied cause-effect link (Saunders et al. 2009, p. 126). As Saunders et al. (2009, p. 126) elaborate with their employee absenteeism example, also in this research it is more realistic to consider the persons who evaluate the service quality as humans, rather than as unthinking research objects responding to circumstances. This is highlighted by the fact that service quality in itself includes very subjective elements.

A division between purely deductive and inductive research is difficult to make (Saunders et al. 2009, p. 127), and it can be argued that also in this research both approaches are utilized. When answering the first research question and the sub-questions derived from it, clearly the research is mostly inductive: the aim is to understand how service quality can be defined and for that purpose qualitative data is gathered (Saunders et al. 2009, p. 127). The latter part of this research is focused on the actual measurement of the supplier service quality, where the developed approach is applied and tested in practice. Quantitative data is gathered through a questionnaire and the relationships between the variables are analyzed. This part can be seen as deductive research. Saunders et al. (2009, p. 127) state that combining both approaches is not only possible, but often also advantageous.

As well as having characteristics from both inductive and deductive research, the purpose of this research is both exploratory and descriptive. Saunders et al. (2009, p. 140) note that this is often the case: descriptive research may be an extension of, as is in this case, or a forerunner to, exploratory research. The importance of the descriptive research stems from the fact that it is necessary to have a clear picture of the researched phenomena. Exploratory research on the other hand is useful if you want to clarify your understanding of a problem. (Saunders et al. 2009, pp. 139-140.)

This research was conducted as a case study. A case study strategy is often used in explanatory and exploratory research (Saunders et al. 2009, p. 146). Voss et al. (2002, p. 195) argue that case research can be very useful to practitioners, and that it is suitable especially for new theory development and theory testing and refinement. Therefore, the adoption of the case research strategy is justified, since the problem and objectives of this research are linked to the case company's current situation. The challenges of conducting case research include time, interviewing skills and the generalizability of conclusions (Voss et al. 2002, p. 195). These became evident during the research process. Different data collection techniques and sources were used in order to validate the research. This is called triangulation, and it is often used in case study research (Saunders et al. 2009, p. 146). Triangulation can strengthen the validity of the research (Voss et al. 2002, p. 208): it is used to "ensure that the data are telling you what you think they are telling you" (Saunders et al. 2009, p. 146). Triangulation was done in several phases of the research. The findings from the literature were strengthened in the interviews, and the interview findings were further confirmed with a confirmatory survey. Documents related to the quality measurement and definition were used in the development of the supplier service

quality measurement. The meetings with the case company representatives offered confirmation and additional insights throughout the research process.

As already mentioned, both qualitative and quantitative data collection techniques and data analysis procedures are needed to answer the research questions of this study. This suggests that multiple methods, and especially mixed-methods approach is used. The research choices of this research are presented in Figure 2. A distinction between qualitative and quantitative data is often made. Saunders et al. (2009, p. 482) emphasize the implications for the analysis of the data. As qualitative data is usually complex and non-standardized, the gathered data likely needs to be summarized, categorized or restructured. Usually the analysis of qualitative data involves the creation of a conceptual framework (Saunders et al. 2009, p. 484), as is the case also in this research. The development of the supplier service quality framework based on qualitative data is discussed in Chapter 4.

Quantitative data on the other hand holds little meaning before it has been processed and analyzed. It is the processing of these data that turns them into information. Quantitative analysis techniques include for example graphs and charts, but also more complex techniques like statistical modelling. (Saunders et al. 2009, p. 414.) In this research, the processing and analyzing of quantitative data is used to interpret the supplier service quality survey results. This is presented in Chapter 5.

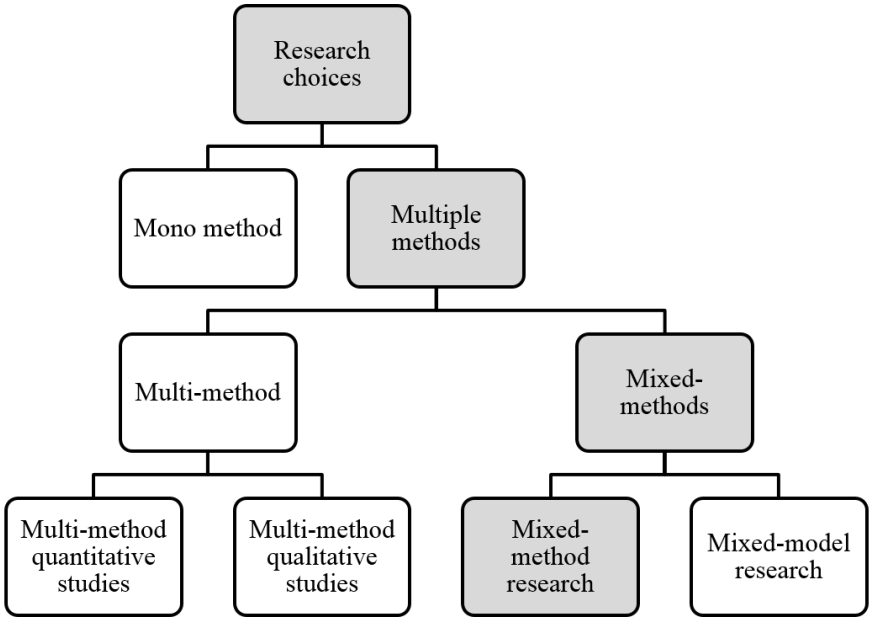


Figure 2. *The research choices of this research (adapted from Saunders et al. 2009, p. 152).*

Within mixed-methods approach, there are two alternatives: mixed-method and mixed-model research. The difference between these two is that in mixed-method research the quantitative and qualitative data collection techniques and analysis procedures are not

combined, i.e. quantitative data are analyzed quantitatively and qualitative data are analyzed qualitatively. (Saunders et al. 2009, pp. 151-153.) Therefore, this research is mixed-method research. The data collection techniques and analysis procedures utilized in this research are discussed in more detail in Sections 5.1 and 6.1.

1.5 Structure of the thesis

The rest of this thesis is structured as follows. In Chapters 2. and 3., the theoretical background and the literature review for this research are presented. First, Chapter 2. presents both purchasing and services briefly, and then combines these two concepts in order to offer a theoretical context for the research. In Chapter 3. service quality, service profit chain and supplier performance measurement literature are reviewed. The emphasis of the chapter is on the previously developed service quality models. At the end of Chapter 3., an initial framework for supplier service quality is presented based on the existing literature.

The framework is used a basis for the empirical part of this research, which is presented in two parts in Chapters 4. and 5. The methodology of the research is discussed in the beginning, and the respective results are presented at the end of these chapters. Chapter 4. describes the development process of the supplier service quality framework based on the interviews and discussions with the case company and the supplier representatives. As a result, Chapter 4. presents the developed supplier service quality framework. This framework is then used a basis for the actual measurement of the supplier service quality. The measurement process along with the supplier service quality results are presented in Chapter 5. Also, the interpretation of the measurement results, as well as the future development of the measurement are discussed. Finally, the conclusions are presented in Chapter 6., including summary of the main findings, managerial implications, limitations and criticism of the research and implications for future research.

2. PURCHASING AND SERVICES

2.1 Practices and classifications of purchasing

In a traditional presentation, the purchasing process can be seen to comprise of six phases: determining specification, selecting supplier, contracting, ordering, expediting and evaluation, and follow-up and evaluation (van Weele 2014, p. 8). This description of the purchasing process is widely used (e.g. Van der Valk & Rozemeijer 2009). The purchasing process is shown in Figure 3.

Figure 3. *The purchasing process (adapted from van Weele 2014, p. 8).*

First, the specifications are determined in terms of quality and quantity. Then the best possible supplier is selected and the contract is drawn and agreed upon. After contracting an order is placed and subsequently monitored and controlled. The last phase is follow-up and evaluation, where the supplier is evaluated. (van Weele 2014, p. 8.) As the subject of this research is supplier service quality, this research is concerned with the end stages of the purchasing process. Supplier service quality measurement is a part of the supplier evaluation, where the supplier has already been chosen and an order has been placed.

Purchasing can be divided into direct and indirect purchasing based on whether the purchased materials and services become directly part of the company's offering or not. Most indirect purchases are services but also direct purchases usually contain some services, depending on the business. (Heikkilä et al. 2013, p. 11.) In purchasing and supply management roughly two opposite forms of purchasing behavior can be identified: transaction-oriented and relationship-oriented behavior (Axelsson & Wynstra 2002, p. 213). Characteristics of both behaviors are shown in Table 1. Transaction-oriented purchasing behavior (also referred to as "classical purchasing philosophy") considers every transaction a new business deal, and benefits are pursued through short-term based competition. Relational approach (also "modern purchasing philosophy") focuses on the interaction between the buyer and the seller, and benefits are pursued through long-term relationships and cooperation. In practice, the same firm can use different approaches for different suppliers, and the two main approaches can also be combined. (Axelsson & Wynstra, 2002, pp. 213-214, 227-233.)

Table 1. *Transaction-oriented and relationship-oriented approaches to purchasing (adapted from Axelsson et al. 2002, p. 54).*

Transactional approach	**Relational approach**
Many alternatives	One or few alternatives
Every deal is a new business deal, no-one should benefit from past performance	A deal is part of a relationship and the relationship is part of a network context
Exploit the potential of short-term based competition	Exploit the potential of co-operation
Short-term; arm's length distance, avoid coming too close	Long-term with tough demands and joint development
Renewal and effectiveness through change of partner, choose the most efficient supplier on each occasion	Renewal and effectiveness through collaboration and "team effects", combine resources and knowledge
Buying "products" ➔ Price-orientation, strong in achieving favorable prices for well-specified products	Buying "capabilities" ➔ Cost- and value-orientation, strong in achieving low total costs of supply and developing new value

The relational approach to purchasing has emerged as a result of the development of the purchasing function (Axelsson & Wynstra 2002, p. 214). The terms classical and modern purchasing philosophy refer to a shift in purchasing philosophy, which Sheth (1996, p. 11) argues is a consequence of the changing paradigms of purchasing. These paradigms include global competitiveness, the emergence of total quality management, industry restructuring and the use of information technologies (Sheth 1996, p. 10). Sheth (1996) describes the shift from transaction-centered to relational-centered purchasing philosophy as a two-dimensional shift, as presented in Figure 4. In addition to the change from transaction centered purchasing to relationship centered purchasing, there is also a shift from domestic sourcing to global sourcing.

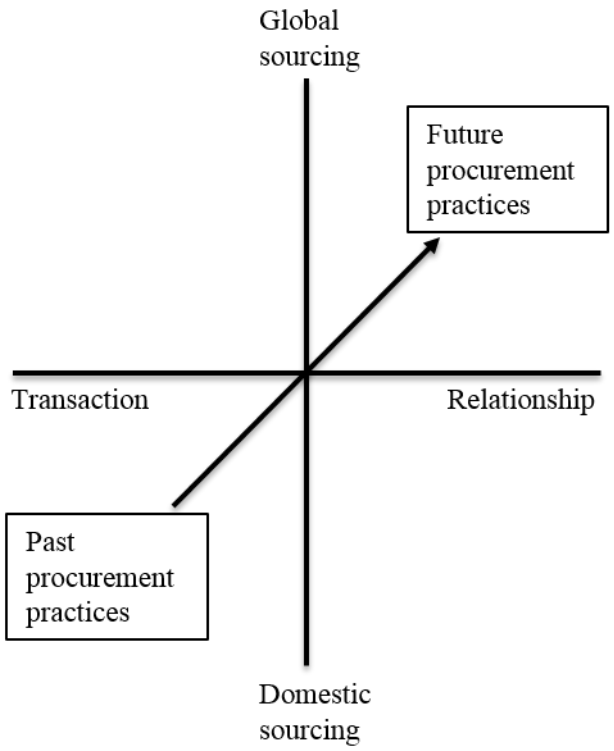


Figure 4. *The two-dimensional shift in purchasing philosophy (adapted from Sheth 1996, p. 11).*

The transactional purchasing model is quite straightforward. First the need is established. Then potential suppliers are approached and compared, and ultimately the best one is chosen. At the last phase the buyer uses competition to get the best possible conditions. The transactional model's reasoning is built on a number of assumptions about the marketplace, including that the demand and supply are well known and matching. The model has some implications on the behavior of the purchasing company. The buying company's purchasing tries to minimize the number of intermediaries and buy directly from the manufacturer in order to lower the price. This results in additional warehousing, administrative and other activities, and also more coordination is needed to handle the increased number of suppliers. Another implication is that independence from suppliers is needed to achieve long-term efficiency. When the buyer becomes dissatisfied with the supplier, it switches suppliers rather than tries to develop the current one. (Axelsson & Wynstra 2002, pp. 214-216.)

The basis of the relational purchasing approach is that sustained competitive advantage is developed through relationships between firms (Dyer & Singh 1998, p. 675). Dyer & Singh (1998, p. 662) argue that that there are four determinants of inter-organizational competitive advantage: relation-specific assets, knowledge-sharing routines, complementary resource endowments and effective governance.

One key difference between the transactional and the relational approach is the attitude towards price: in the transactional view price is the main driver, whereas in the relational

view the focus is on achieving low total costs. Therefore, the relational view takes into account all the steps (and the associated costs) from the production of a product to the use within the buying company. The relation-oriented view of purchasing emphasizes combining the resources and knowledge of both the buyer and the supplier. In order to achieve this, a long-term commitment is usually required. (Axelsson & Wynstra 2002, pp. 213-217.)

Whether the transactional or relational purchasing approach is more suitable, depends on the specific situation. Usually the relational approach is better when the indirect costs are substantial, because then also other costs are relevant in addition to price. On the other hand, when the direct costs are large, the transactional approach might be better. (Axelsson & Wynstra 2002, p. 219) However, Axelsson & Wynstra (2002, p. 219) suggest that nowadays the relational approach seems to be the right choice more often than before. Zimmermann & Foerstl (2014, p. 47) found in their study, that supplier-facing relational practices have larger effects on buying firm performance than non-relational purchasing practices.

It can be argued therefore that the relational purchasing approach should be used with key suppliers, when the purchased product or service has a substantial effect on either the firm or the end-product, or when the service is otherwise important for the buying firm. In that case, pursuing low total costs and developing common practices are worth the effort.

2.2 Services and their characteristics

Services are a complex phenomenon. Several different definitions for services exist, and there is not one that is universally agreed upon. (Grönroos 2007, p. 51-52.) Grönroos (2000, p. 46) offers the following definition for services:

"A service is a process consisting of a series of more or less intangible activities that normally, but not necessarily always, take place in interactions between the customer and service employees and/or physical resources or goods and/or systems of the service provider, which are provided as solutions to customer problems."

This definition emphasizes the interactions involved in services, which are an important part of services (Grönroos 2000, p. 46). Another definition of services is suggested by Vargo & Lusch (2004a, p. 2), who define services as "the application of specialized competences (knowledge and skills) through deeds, processes, and performances for the benefit of another entity or the entity itself". Vargo & Lusch (2004a, p. 2) argue that while compatible with Grönroos' (2000) definition, their definition is more inclusive. Vargo & Lusch's (2004a) definition for services is very broad, which represents well the complex nature of services.

In the literature, services have been traditionally seen as different from goods, or as "what tangible goods are not" (Vargo & Lusch 2004b, p. 325). The four most commonly mentioned differences between goods and services are the so called IHIP characteristics: intangibility, heterogeneity, inseparability and perishability (e.g. Zeithaml et al. 1985). Intangibility refers to the fact that services can't be touched. Heterogeneity means that every service is unique, due to the fact that services include people and every person is unique. This means that service exchanges cannot be standardized. (van Weele 2014, p. 79.) Inseparability refers to the simultaneous production and consumption of services. Perishability means that services can't be stored. (Zeithaml et al. 1985, pp. 33-34.) These characteristics have also received criticism (e.g. Vargo & Lusch 2004b; Edvarsson et al. 2005).

Vargo & Lusch (2004b) argue that the IHIP characteristics don't distinguish services from goods, and that the IHIP characteristics only have meaning in a manufacturing context. In addition, they argue that a service can be provided directly, or indirectly through tangible goods, which entails that "goods are distribution mechanisms for service provision" (Vargo & Lusch 2004b, p. 326). It has been suggested, that a pure service or pure material seldom exists (Smeltzer & Ogden 2002, p. 58): the exchange of goods contains some form of service and many services contain also tangible goods (Iloranta & Pajunen-Muhonen 2015, p. 208). Edvarsson et al. (2005, p. 118) on the other hand argue that definitions of service are too narrow, and that the service characteristics (including IHIP characteristics) are outdated.

Lovelock & Gummeson (2004) point out, that many services actually possess also the opposite characteristics of the IHIP characteristic, i.e. tangibility, homogeneity, separability and durability. They argue, that there are tangible processes and outcomes involved with some services: customers feel, see and can observe physical outcomes. There are also tangible impacts to customers' possessions in, for example, cleaning services. The reliability and consistency of delivering some services (e.g. freight transport and dry cleaning) has been significantly improved through improvements in service quality and automation. Especially many services directed at physical possessions (e.g. warehousing and repair) do not require customer presence in the actual production process. (Lovelock & Gummeson 2004, p. 31) However, in some services the customer still is present for certain parts of the process, for example, when bringing the car to the repair shop and paying (Grönroos 2007, p. 52; Lovelock & Gummeson, p. 31).

Durability can be detected especially in education, entertainment and news industries, where service performances and output can be recorded (Lovelock & Gummeson 2004, p. 31). Lovelock & Gummeson (2004, p. 32) suggest that changes in the service industry have affected the validity of the IHIP characteristics. These changes include increased automation, quality improvement procedures, outsourcing and the development of information technology. The focus in service research has shifted away from defining services, since an agreed definition might well be unnecessary or even impossible: more

could be accomplished by examining the common characteristics of services and the nature of service consumption (Grönroos 2000, p. 47).

Grönroos (2007, p. 53) identifies three common general characteristics of services:

1. Services are processes that consist of activities or a series of activities.
2. Services are at least to some extent produced and consumed simultaneously.
3. Services are at least to some extent co-produced with the customer.

These characteristics have strong implications especially on the quality control and measurement of a service. First, due to the simultaneous production and consumption (i.e. inseparability), there is no perceptible quality before the service is produced and consumed. Therefore, the measurement of service quality must take place at the time and place of the service delivery. Due to the intangible nature, the evaluation of the service is often difficult for the customer. It is important to notice, that some parts of the service process may not be visible to the customer: in those cases, only the outcome is experienced. An example of this is the delivery of goods, where most of the service process is invisible to the customer. The measurement of service quality should of course focus on those aspects of the service that are visible to the customer. Furthermore, due to the people, personnel and customers involved in the service delivery process the delivered service is never exactly the same. This inconsistency makes it challenging to maintain an even service quality. (Grönroos 2007, pp. 54-55.) It also further emphasizes the importance of service quality measurement as a way of managing the service production and delivery process.

2.3 Classification of services

The purposes of developing service classification schemes are many. The first typologies concentrated on the definition of services. In the 1960s services were distinguished from manufactured goods, and service marketing became the main purpose for classifications. The purposes of classification also include strategy and managerial issues. However, most service classifications have been developed in relation to the research context and since offer a quite narrow view on service issues in general. (Cook et al. 1999, pp. 321-322.)

There are several different ways to classify services (for an overview, see for example Cook et al. 1999). The term “business services” is defined in this research according to Heikkilä et al. (2013, p. 26) as “all types of professional services purchased by business organizations”. Heikkilä et al. (2013, p. 26) divide business services further into knowledge-intensive business services (KIBS) and industrial services. In this research Heikkilä et al.’s (2013) definition for industrial services will be used. Industrial services “support the utilization of productive assets and operative processes, such as production and transportation equipment and their use” (Heikkilä et al. 2013, p. 129). Industrial

services include for example the maintenance of production equipment. Cleaning services, from which the data in this research will be gathered, is also an industrial service.

Van der Valk (2007) suggests (based on Wynstra et al. 2006) the classification of services into four service types based on the actual use of the service by the customer organization. The four categories are component services, semi-manufactured services, instrumental services and consumption services. The cleaning of production facilities is a consumption service. Van der Valk (2007) has also identified different patterns of interaction for these service types. For consumption services the objective is to ensure that the service supports core organizational processes. In the buyer's side, internal customers are strongly involved in the interactions and substantial administrative efforts are present. Also translating internal customer demands is an important buyer capability. The focus from the supplier's side is on developing efficient routines. Communication concerns daily activities and service performance. (Van der Valk 2007, p. 297; Van der Valk et al. 2009, p. 825.)

To overcome the problem of industry specificity, Lovelock (1983) suggests a classification of services into discretely and continuously delivered services, based on the nature of the relationship with the customer. Grönroos (2007, p. 57) emphasizes this classification and argues, that the continuous flow of interactions between the customer and service provider in continuously rendered services (e.g. cleaning and security services) creates an opportunity to develop valued relationships with the customers. This is often more difficult for the providers of discretely used services (e.g. ad hoc repair services), even though still possible and arguably profitable. For the providers of continuously used services the cost of finding new customers is often high, and hence they cannot afford to lose customers. (Grönroos 2007, p. 57.) Many of the services purchased in b2b context are continuously delivered services (Lovelock 1983, p. 13). While offering implications concerning for example pricing (Lovelock 1983, p. 14), this in part provides motivation for the supplier of continuously delivered services to develop and strengthen the relationship with the customer. A mutual interest is of vital importance when developing and implementing a common measurement of any kind between two companies. The supplier's motivation for the common service quality measurement is discussed more in Section 4.

Chase (1978) suggest a classification of services into high and low contact services, based on the extent of required customer contact in the creation of a service. Customer contact is defined as the physical presence of the customer, and creation of the service refers to the work process in which the service is produced. This classification highlights the amount of interaction between the supplier and the customer. In low-contact services the quality standards are usually measurable. A high-contact service results in subjective, and thus variable, quality assessment and control. (Chase 1978, p. 138-139.) Furthermore, Chase (1978, p. 140) notes that "Any interaction with the customer makes the direct

worker in fact part of the product and therefore his attitude can affect the customer's view of the service provided".

It is important to understand the effect of the purchased services on the organization and its processes. Even a low-volume service might be critical, if the service is linked to the core competence of the organization. (Iloranta & Pajunen-Muhonen 2015, p. 203.) Van Weele (2014, p. 81) states that the way companies classify services has implications on purchasing decision-making and on the operational phase after contract closure. Fitzsimmons et al. (1998, p. 378) conclude that classification of services facilitates the purchasing process, since a similar approach can be used to other services in the same category. Grönroos (2007, p. 57) states that although understanding classifications of services is important for (the service provider's) management, all services are in some way unique and this should always be taken into account. Similarly, Smeltzer & Ogden's (2002, p. 66) findings suggest that all services should not be considered as one class of purchases, since there are differences within the purchasing processes.

2.4 Purchasing services

The service purchasing spend is still small compared to the materials purchasing spend, but the share of services is constantly growing (Heikkilä et al. 2013, p. 11). In addition to pure services purchased, many of the tangible items purchased (e.g. materials and components) include at least some form of service (Iloranta & Pajunen-Muhonen 2015, p. 202).

Although not unambiguous, the differences between goods and services pose some challenges for the purchasing of services. The research concerning service purchasing is very limited compared to the product purchasing literature (Carter & Elram 2003). However, it is suggested that the purchasing process for services and materials should not be the same (Smeltzer & Ogden 2002, p. 67). Van Weele (2014, p. 78) states that it is more difficult to go through the different stages of the purchasing process when buying services. Smeltzer & Ogden (2002, p. 67) studied perceived differences between purchasing materials and purchasing services using focus groups and interviews of purchasing professionals. They found that purchasing professionals perceive service purchasing as more complex than buying materials, whereas top management perceives service purchasing as less complex. Fitzsimmons et al. (1998, p. 372) say that the purchasing process of business services is more complicated than in case of material goods because of the larger number of affected personnel. For example, food services affect all employees personally, whereas materials acquired for the production process do not.

Van der Valk & Rozemeijer (2009, p. 5-6) studied the perceptions of Dutch purchasing managers on the differences between buying goods and services. The research was

conducted using a survey, to which they got 71 responses. Based on the survey findings, the researchers identified three problem areas in the purchasing process for services:

1. Specifying the service
2. Defining the specific content of a service level agreement
3. Evaluating performance

The specification problem is caused by the fact that it is difficult to identify the content of the service before it is purchased (Van der Valk & Rozemeijer 2009, p. 6). This relates to the intangibility and the simultaneous production and consumption of services (Axelsson & Wynstra 2002, p. 139). Jackson et al. (1995, p. 104) found that buyers perceive determining specifications for services more difficult than for goods. Iloranta & Pajunen-Muhonen (2015, pp. 209-210) state that the specification of goods is easier, since one can physically measure the dimensions of the product. The specification of services is usually not so straightforward, and even accurate descriptions of services may end up containing subjective measures and dimensions. When the specification phase is not done thoroughly, it will be problematic to define the specific content of a service level agreement (Van der Valk & Rozemeijer 2009, p. 6). Service level agreement (SLA) is a contract that describes the performance required from the supplier (van Weele 2014, p. 79). Insufficient specification and SLA may lead to a situation where neither the buyer nor the supplier knows what actually should be measured and how. This in turn makes the performance evaluation of the supplier and the service difficult. (Van der Valk & Rozemeijer 2009, p. 6.) Also Smetltzer & Ogden (2002, p. 67) report that in their study, evaluating supplier performance was rated as the most complex step in the purchasing process for services.

To overcome the difficulties, Van der Valk & Rozemeijer (2009, pp. 6-7) suggest that two additional steps should be incorporated into the purchasing process of Van Weele (2014) for services: Request for information and detailed specification. The resulting purchasing process for services is presented in Figure X.

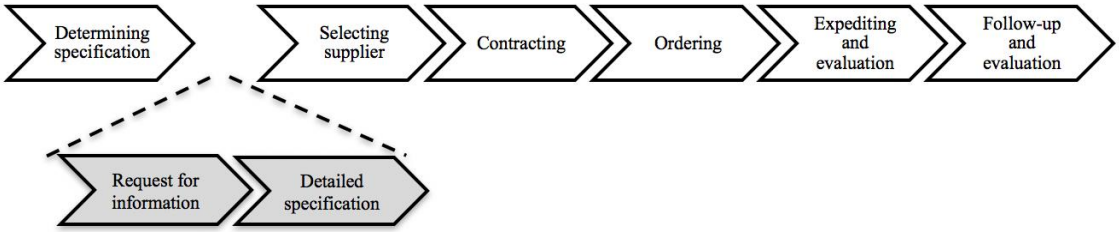


Figure 5. *The service purchasing process suggested by Van der Valk & Rozemeijer (adapted from Van der Valk & Rozemeijer 2009, p. 7).*

Van der Valk & Rozemeijer's (2009, p. 6-7) purchasing process for services highlights the importance of the interaction between the buyer and the potential sellers. By

requesting information from the suppliers at an early phase of the process, the buyer is able to compare different solutions, and the gathered information can be used to develop a detailed specification. In addition, the internal client should be included in the determination of the SLAs. The two added steps propose that the service should be developed jointly to satisfy the requirements of both parties, to which Van der Valk & Rozemeijer (2009, p. 7) refer to as "service development".

In this thesis, the performance evaluation of the supplier and the service is addressed. For that purpose, the service quality of industrial services will first be defined to ensure that the evaluation is targeted at relevant aspects of the service. In this way, the performance evaluation will also advance the content specification of the service level agreement.

For specifying business services, Axelsson & Wynstra (2002, p. 144) suggest four possible methods: Input, throughput, output and outcome. In input-oriented service specification the focus is on the resources and capabilities of the supplier: The resources and possibly the quality of the resources that should be spent by the supplier are defined. An example of this is buying consultancy: the problem or the solution is not known, but it is expected that a consultant is able to solve it. In this case the actual service bought is very loosely defined. Throughput-oriented service specification is concerned with the process of the service: the customer defines the activities it wants to be performed in order to achieve the wanted outcome. This requires that the customer knows what it needs and can precisely describe the activities to be performed. The expectation here is that when the supplier performs the described activities the wanted outcome is achieved. In output-oriented service specification the focus is on the function or performance of the service: The customer defines the output and lets the supplier choose the appropriate activities. It is also possible that the customer and supplier together define the activities and resources as well as the desired output. The outcome-oriented service specification focuses on the value for the user: The customer defines what the service should accomplish. (Axelsson & Wynstra 2002, p. 143-148.)

The case company of this research has chosen to specify the cleaning service mostly according to the output-oriented service specification. The case company has quite precisely defined the desired output of the cleaning service and partly also the activities it wants to be performed. In this case, the description of output is very technically detailed, and the output of the cleaning service is monitored with monthly quality rounds. Therefore, it can be argued that the output-oriented service specification facilitates the evaluation of the service, since measures for service quality can be chosen based on the output specification of the service. However, due to the technical nature of the output specification, also the measures used to evaluate cleaning service quality are all technical (i.e. objective). In output-oriented specification the buyer must ensure that the specified output actually leads to a wanted outcome, i.e. good service quality. The sole use of objective measures can be considered problematic, because objective measures alone do not guarantee good service quality; they merely indicate whether something is as it had

been specified. Therefore, the evaluation of the service requires measures that indicate whether the service has been successful or not.

3. SERVICE QUALITY AND SERVICE PROFIT CHAIN

3.1 Service quality

3.1.1 Quality

Quality can be described as an indistinct and elusive construct, as everyone knows the meaning of quality for oneself (Marquardt et al. 2017, p. 96). Countless definitions for quality exist, which indicates that the term "quality" is largely dependent on the context in which it is used. The academic literature on quality is extensive: quality has been explored in philosophy, economics, marketing and operations management literature. Each of these disciplines have had a differing point of view, resulting in competing perspectives on quality. (Garvin 1988, p. 39.) For the purposes of this research, the views of marketing and operations management on quality will be further examined.

In marketing, a product-based or user-based approach to quality is often adopted. Product-based approach views quality as a precise variable that can be measured: quality is reflected in, for example, an attribute of a product. Therefore, the quality of an engine could be measured in terms of horsepower. The user-based approach, on the other hand, incorporates the needs and wants of the consumer to the definition of quality: the highest quality goods are those that best satisfy the customer's preferences. A clear difference between product-based and user-based approaches exist: product-based approach treats quality objectively, whereas in user-based approach quality is highly subjective. (Garvin 1988, pp. 42-43.)

Both approaches have their own strengths and weaknesses. The objective nature of quality means that it can be easily measured and assessed. However, a higher quality can only be achieved at higher costs, as increasing attributes in a product is considered to be costly. Another weakness of the product-based quality approach is that a correspondence between product attributes and quality does not always exist in practice: the quality of a product can actually be based on something else entirely. The user-based approach's strength lies in observing customer preferences, as the customer ultimately decides the quality of the product. However, the subjectivity raises also a problem, as the varying individual preferences complicate the defining of quality in an aggregated level. Another problem is with the maximization of customer satisfaction: is a product that maximizes satisfaction necessarily better? (Garvin 1988, pp. 42-44.)

In operations management, the focus is on engineering practices and manufacturing control. Hence, a manufacturing-based approach to quality is adopted: quality is seen as

conformance to requirements. Good quality is achieved when the product or service meets specifications. On one hand, quality is pursued through reliability engineering, and on the other hand, through statistical quality control. Both of these techniques aim for cost reduction: improvements in quality lead to lower costs due to a decrease in repairs and rework. A weakness of the manufacturing-based approach to quality is that the focus is mostly internal: no consideration is given to the customer's preferences. (Garvin 1988, pp. 44-45.)

As Garvin (1988, p. 47) states, understanding different perspectives on quality can be beneficial to companies. Moreover, using only one approach to quality is likely to cause problems. This is easy to understand: a good conformance to requirements does not in itself guarantee that the product is considered to be of good quality by the customers. Similarly, even if the product has all the features that customers want, but the product unintentionally breaks down after a couple times of use, or half of the manufactured products need to be scrapped, the product probably would not generally be considered to be of good quality.

Traditionally, quality has been seen as conformance to specification (e.g. number of defects produced, cost of quality) (Neely et al. 1995, p. 84), as quality has been first considered in the context of goods (Parasuraman et al. 1985, p. 42). However, with total quality management (TQM) the focus has shifted on customer satisfaction (Neely et al. 1995, p. 85). Generally, quality and customer satisfaction are seen as distinct constructs, yet highly correlated (Dabholkar et al. 2000, p. 166). Customer satisfaction is closely related to the subjective user-based quality approach (Garvin 1988, p. 44), and therefore also closer to the subjective quality. Emphasizing the customer in defining quality makes sense, because in practice the product or service cannot be considered as being of good quality, if it is not in accordance with the customer requirements. Only after finding out the customer preferences should the company pay attention to other perspectives of quality. Grönroos (1988, p. 11) notes that organizations must define quality in the same way customers do, or otherwise any actions taken to improve quality may be meaningless: "it is quality as it is perceived by the customers that counts".

The evaluation of service quality is considered to be more difficult compared to goods quality, as in the case of services there are fewer tangible cues on which to evaluate the quality (due to the intangibility of services), and therefore the evaluation is based on other cues (Parasuraman et al. 1985, p. 42). Since in services the production and consumption take place simultaneously, there are frequent interactions between the customer and the service provider. These interactions actually have a large impact on the perceived service, and therefore also on the service quality perceived by customers. (Grönroos 1988, p. 11.)

3.1.2 Service quality models and dimensions of service quality

Service quality has received a lot of attention in the academic literature. Numerous models have been developed for service quality and service quality measurement. Some of the models have been built to measure the service quality of a certain service or industry (e.g. Dabholkar et al. 1996), while others try to capture the more general dimensions of service quality (e.g. Parasuraman et al. 1985). Nonetheless, it seems that an agreed upon general definition and model for service quality measurement does not yet exist (Seth et al. 2005, pp. 933-934). The best-known service quality models are probably the ones by Grönroos (1982) and Parasuraman, Zeithaml and Berry (1985).

Grönroos (1982, 1984) originally defined service quality as a concept of perceived service quality. Perceived service quality is the result of an evaluation process, where the customer compares the perceived service and the expected service. Later on the model would take the form described in Figure 6. (Grönroos 2007, p. 77.)

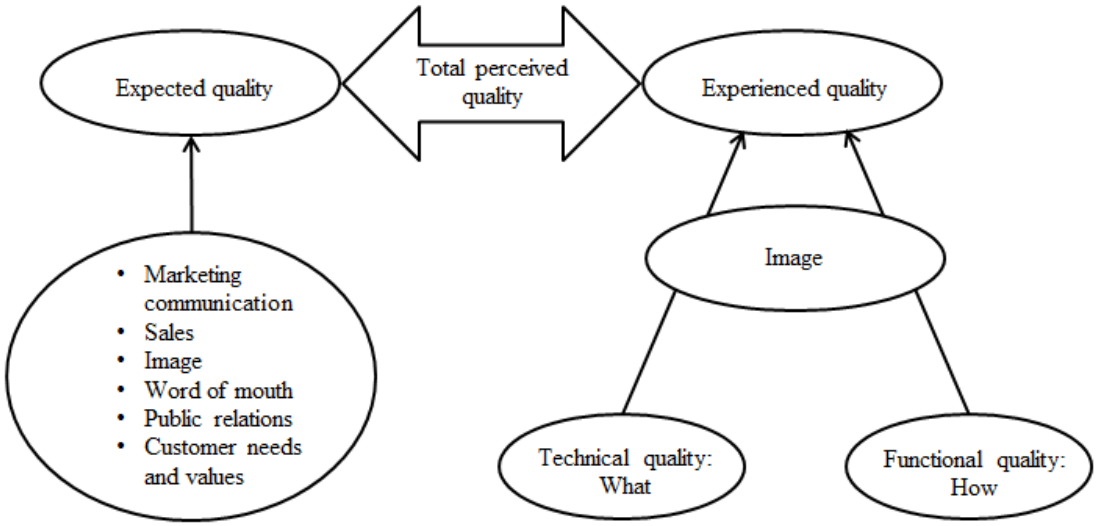


Figure 6. *Grönroos' service quality model (adapted from Grönroos 2007, p. 77).*

Two dimensions affect the experienced quality: technical and functional quality, i.e. what the customer gets, and how he gets it. Technical quality represents what the customer receives in the interactions with the firm, i.e. the technical quality of the outcome of the service. Technical quality is what the customer is left with after the service production process. Functional quality describes the quality of the process in which the outcome is produced, i.e. the functional quality of the process. The functional quality includes, for example, the interactions between the service provider and the customer. This means that the appearance and behavior of the service personnel, i.e. what they say and do during the service production process, has a significant impact on the experienced quality. The company image can also affect the experienced quality and it can be viewed as filter: for

example, if the company has a favorable image, it is likely that minor mistakes will be forgiven. The technical quality dimension can often, but not always, be measured relatively objectively, since it is basically a technical solution to a problem. The functional quality on the other hand cannot be evaluated as objectively, since it is perceived very subjectively. (Grönroos 2007, pp. 73-74.)

The experienced quality is compared to the expected quality, and the total perceived quality is obtained as a result. The expected quality is affected by a number of factors, e.g. marketing communication, image and word of mouth. The service quality is good, when the experienced quality meets the expected quality, i.e. the customer expectations are met. It is also important to notice, that if the expected quality is at an unrealistic level, the total perceived quality will be low, even if the experienced quality is good. The level of total perceived quality is then determined by the gap between the expected and experienced quality. Therefore, Grönroos' model implies that when considering service quality, both the operative and marketing aspects need to be taken into account. (Grönroos 2007, pp. 76-77.)

Servqual was introduced in 1988 by Parasuraman, Zeithaml and Berry, although the idea behind the model was already published in the article "A conceptual model of service quality and its implications for future research" in 1985 by the same authors. Servqual is based on Grönroos' (1982) notion that quality is a comparison between expectations and performance (i.e. the disconfirmation paradigm). Parasuraman et al. (1985) proposed that service quality can be modeled using gaps, i.e. discrepancies in executive perceptions of service quality and the actual tasks designed to deliver the service to consumers. Four of the five gaps in the model are in the service provider's side and measure these discrepancies. The fifth gap measures the difference between the consumer's expectation and perception of the service. (Parasuraman et al. 1985, p. 46.) The Servqual model with the gaps is presented in Figure 7.

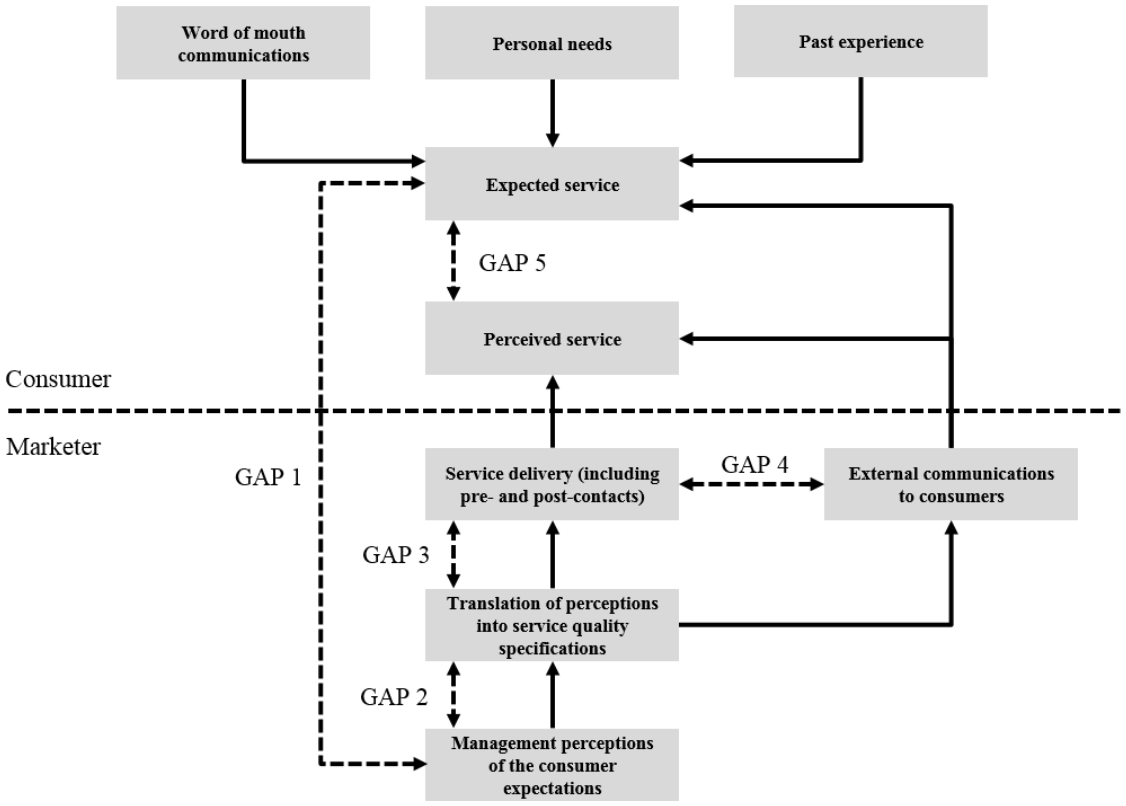


Figure 7. *The Servqual service quality model (adapted from Parasuraman et al. 1985, p. 44).*

Parasuraman et al. (1985) used executive and focus group interviews in their exploratory study. Originally, they recognized 10 determinants that consumers use to assess service quality, but later they reduced the number of determinants to five: tangibles, reliability, responsiveness, assurance and empathy (Parasuraman et al. 1988). The first three determinants were also in the original 10, but the last two, assurance and empathy, contain items from seven original dimensions: competence, access, courtesy, communication, credibility, security and understanding the customer. Parasuraman et al. (1988, p. 24) state that the five dimensions still capture the aspects of all the 10 original dimensions. In 1991, Parasuraman et al. (1991) revised the model again, but the five dimensions stayed intact. The definitions of the five dimensions are presented in Table 2. The five dimensions are measured with a total of 22 items: the items have both an expectation and a perception counterpart, so the comparison between expectations and perceptions is possible.

Table 2. *The Servqual dimensions and their definitions (adapted from Parasuraman et al. 1988, p. 23).*

Dimension	Definition
Tangibles	Physical facilities, equipment, and appearance of personnel
Reliability	Ability to perform the promised service dependably and accurately
Responsiveness	Willingness to help customers and provide prompt service
Assurance	Knowledge and courtesy of employees and their ability to inspire trust and confidence
Empathy	Caring, individualized attention the firm provides its customers

Servqual has received a lot of criticism. The use of expectations versus perception in the measurement of service quality has been questioned (e.g. Cronin & Taylor 1992; Asubonteng et al. 1996). Cronin & Taylor's (1992, pp. 63-64) findings suggest that service quality should be measured as an attitude, and that measuring only perceptions (performance) is a better indicator of service quality. They developed a performance-only measurement of service quality called Servperf, using the perception items from Servqual. The performance-only measurement has been found better also in other studies (Dabholkar et al. 2000; Durvasula et al. 1999; Brady et al. 2002). Based on the service quality literature and his findings, Teas (1993, p. 29) points out that the Servqual's expectations concept is unclear, and he questions the theoretical justification for using the service expectation concept in the measurement of perceived service quality.

Ladhari (2008, p. 76) argues that the Servqual model is actually based on process quality rather than outcome quality. Brady & Cronin (2001, p. 34) also suggests that Servqual model emphasizes the process quality of the service, noting that Servqual "uses terms that describe service encounter characteristics". This is rather easy to see when looking at the definitions of the Servqual dimensions in Table X. Practically all the dimensions describe characteristics and behavioral aspects of the personnel of the service provider, while the actual outcome of the service is not addressed at all. Brady & Cronin (2001, p. 37) argue about the Servqual's dimensions that although there are many aspects in a service that should be, for example, reliable, reliability in itself is not a clear dimension. Therefore, it should be clearly defined what needs to be reliable.

Many of the subsequently developed service quality models are at least partly based on Grönroos' and Parasuraman, Zeithaml and Berrys's work. Holmlund & Kock (1995) suggest economic quality as a third dimension for service quality in buyer-supplier relationships in addition to functional and technical quality. According to the researchers, the economic quality dimension "implies for the buyer that the relationship has to be

profitable" and to the supplier "that the received price covers total costs, including both internal and external quality failure costs" (Holmlund & Kock 1995, p. 118). Based on the preceding research on Servqual, Dabholkar et al. (1996) proposed a hierarchical structure to the measurement of service quality in retail stores. They chose the service quality dimensions based on a literature review and qualitative research. The dimensions were: physical aspects, reliability, personal interaction, problem solving and policy.

Following the work of Dabholkar et al. (1996), Brady & Cronin (2001) developed a hierarchical service quality model that describes service quality as a third-order factor model. This model is presented in Figure 8. In their model, service quality consists of three primary dimensions: interaction quality, physical environment quality and outcome quality. Each of the three dimensions has also three subdimensions, which in turn consist of three individual items. Brady & Cronin (2001, p. 38) didn't identify the Servqual dimensions as direct determinants of service quality; rather, they used the reliability, responsiveness and empathy variables as descriptors of the nine subdimensions in their model.

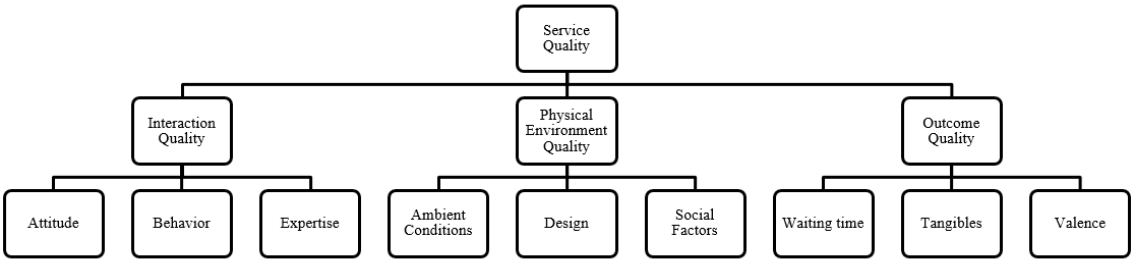


Figure 8. *The hierarchical model of service quality proposed by Brady & Cronin (adapted from Brady & Cronin 2001, p. 37).*

Gounaris (2005) is one of the few to develop a service quality measure for business-to-business services. He developed a scale called Indserv, in which he also adopted a hierarchical approach. His findings suggest that service quality in b2b context is evaluated by four subdimensions: potential quality, hard quality, soft quality and output quality. Potential quality represents "attributes that organizational customers need to evaluate and consider in advance of the provision of the service". Potential quality aims to respond to the uncertainty there is about the performance of the service due to increased complexity and customization of b2b services. Hard quality refers to the suitability of the service delivery processes to produce the wanted outcome. Soft quality is the evaluation of interactions between the supplier's first line personnel and the customer. Output quality captures both the actual outcome of the service and the impact of the delivered service to the buyer. This is based on Gounaris' (2005) findings that organizational customers don't make a distinction between the outcome of the service and the effects of the service to their business. (Gounaris 2005, p. 427; 430.)

The numerous service quality models reflect well the complex nature of services and further support the notion that a generally applicable service quality definition or model does not exist. Ongoing debate exists in the service quality literature concerning, among other things, the dimensions and structure of service quality (Ladhari 2008, p. 78). Despite the received criticism, the Servqual dimensions and the perception-minus-expectations score has been widely used in the literature (Ladhari 2008, p. 75). Servqual seems to be one of the most popular tools for service quality measurement for researchers (Seth et al. 2006, p. 449). Especially, even though Servqual was originally developed using consumer services, majority of the studies in business-to-business context employ the Servqual instrument (Gounaris 2005, p. 422). The reasons behind the popularity of Servqual might be that it is well documented and that it identifies directions for service quality improvement (Seth et al. 2006, p. 449). Ladhari (2008, p. 79) states that the Servqual scale "continues to be the most useful model for measuring service quality". Therefore, the Servqual model of service quality cannot be bypassed when talking about service quality. However, due to its focus on process quality, the ambiguous dimension definitions and the criticized disconfirmation measurement, it was not deemed suitable considering the objective of this research.

For the purposes of this research, a more detailed analysis on the dimensions of service quality is needed. The most cited service quality dimensions are the ones identified by Parasuraman, Zeithaml and Berry (1985, 1988) and Grönroos (1982). However, Brady & Cronin (2001, p. 34) state that perceptions of service quality are clearly based on multiple dimensions, but that there is no consensus about the nature or content of these dimensions. Seth et al. (2006, p. 452) note that "It appears that there is no universal set of dimensions for measurement of service quality; rather they need to be reviewed in the light of a specific service encounter". Therefore, 15 service quality models and the respective dimensions from academic literature were reviewed. The reviewed models and their dimensions are presented in Appendix H. The models were chosen based on their established use in the academic literature and relevance. From Appendix H it can be seen, that there is significant variation in the dimensions of service quality.

In the reviewed service quality models, the number of dimensions varies between two (Lehtinen & Lehtinen 1991) and five (e.g. Ko & Pastore 2005). Two dimensions of service quality are suggested by Lehtinen & Lehtinen (1991, p. 288; 291), who suggest that process and output quality view the service quality more from the customer's point of view than their alternative three dimensions, i.e. physical quality, interactive quality and corporate quality, which are linked to the service organization. The process and outcome quality are very similar to Grönroos' (1984) functional and technical quality. Furthermore, Grönroos' quality dimensions are clearly visible in many of the models in Appendix H. Process and/or outcome quality are present in several of the models (e.g. Gounaris 2005; Brady & Cronin 2001; Collier & Bienstock 2006; Homburg & Garbe

2001). In addition, there are dimensions in the models that can be clearly linked to the process quality (e.g. personal interaction, people's behavior).

From Servqual's dimensions mainly reliability is found in the reviewed models (Cronin & Taylor 1992; Dabholkar et al 1996; Dabholkar 2000). These findings are of course affected by the limited number of service quality models reviewed here. The importance and popularity of the Servqual scale can be seen in many service quality models (Seth et al. 2005, p. 934). For example, Ladhari (2008, p. 68; 79) reviewed 30 developed measures for service quality from 1990 to 2007, and found that the Servqual dimensions (especially tangibles and empathy) were retained in many of the models. According to Ladhari (2008, p. 79), these similarities suggest that some service quality dimensions are generic and some are dependent on industry and context. He also found that the dimensions of the models varied even within the same industry (Ladhari 2008, p. 78).

The multidimensionality of the models is evident also in Table X. From the reviewed models, only the model of Dabholkar et al. (2000) defines service quality as a unidimensional construct. Also Seth et al. (2006, p. 456) found evidence of this. The structure suggested by Parasuraman et al. (1988) (tangibles, reliability, responsiveness, assurance and empathy) has been found unsuitable in many instances, both in b2b and b2c contexts (e.g. Carman 1990; Cronin & Taylor 1992; Dabholkar et al. 1996; Durvasula et al. 1999; Gounaris 2005). Therefore, it is understandable that the reviewed models have a different dimensional structure than Servqual. Brown et al. (1993, p. 138) note that "it takes more than the simple adaptation of the SERVQUAL items to effectively address service quality in some situations". Grönroos (2007, p. 86) states that the Servqual dimensions offer a great starting point for when trying to understand the characteristics of any given service, even though the dimensions should always be customized to describe the specific service.

3.2 Service profit chain

Linkage research links the perceptions of employees and customers in order to explain the antecedents of customer outcomes (e.g. customer satisfaction) (Bowen 2008, p. 157). One of the first studies of linkage research was conducted by Schneider et al. (1980), who found a link between employee perceptions of organization's practices and customer perceptions of service quality in banks. They measured employees' perceptions using variables such as job satisfaction, enthusiast orientation and effort rewarded. Linkage research connects the management practices designed to enhance employee service capabilities and the performance of a firm. Furthermore, it directs the attention to the internal practices that have the most effect on organizational outcomes. (Pugh et al. 2002, p. 82.) The organizational outcomes referred to in the literature include for example customer satisfaction (Schneider et al. 2005), customer loyalty (Salanova et al. 2005), sales (Gelade et al. 2005) and profitability (Anderson & Mittal 2000).

Linkage research has mostly focused on business-to-consumer context (Fischer 2012, p. 60). Linkage research is multidisciplinary, (Bowen 2008, p. 157) which can be seen as a result from combining perceptions, satisfaction and profitability measurements. The multidisciplinary nature is evident: the journals publishing articles related to linkage research include for example Journal of Marketing, Academy of Management Journal, Journal of Retailing, Journal of Applied Psychology and Journal of Occupational and Organizational Psychology.

Service profit chain (SPC) is a modelling of the linkage research (Bowen 2008, p. 164). Service profit chain research links together the operations of a service organization, employee satisfaction, customer loyalty and organizational outcomes such as profit and growth. SPC is aimed at helping organizations to develop their service and the satisfaction of their customers. (Heskett et al. 1994, pp. 164-165.) The depiction of the service profit chain by Heskett et al. (1994) is presented in Figure 9.

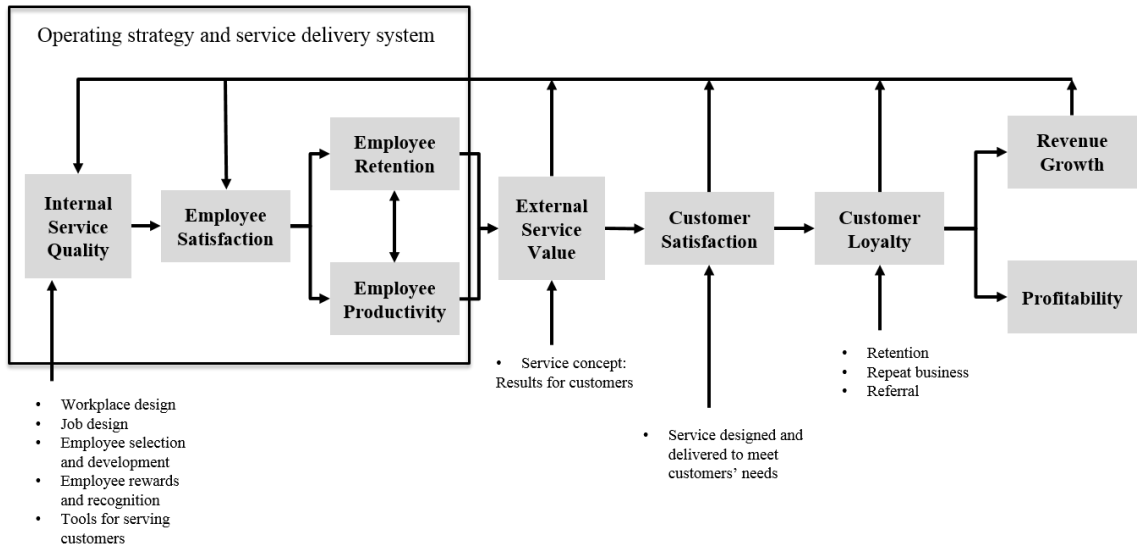


Figure 9. *The service profit chain (adapted from Heskett et al. 1994, p. 166).*

Figure 9. also presents the links in the SPC suggested by Heskett et al. (1994). The seven distinct links that Heskett et al. (1994, pp. 165-169) use to explain the service profit chain are:

1. Internal quality drives employee satisfaction
2. Employee satisfaction drives loyalty
3. Employee loyalty drives productivity
4. Employee productivity drives value
5. Value drives customer satisfaction
6. Customer satisfaction drives customer loyalty
7. Customer loyalty drives profitability and growth

Heskett et al. (1994) developed the model analyzing successful service organizations, but they didn't provide statistical evidence on the existence of these links. A general model based on the existing linkage research literature is depicted by Bowen (2008). This model is presented in Figure 10.

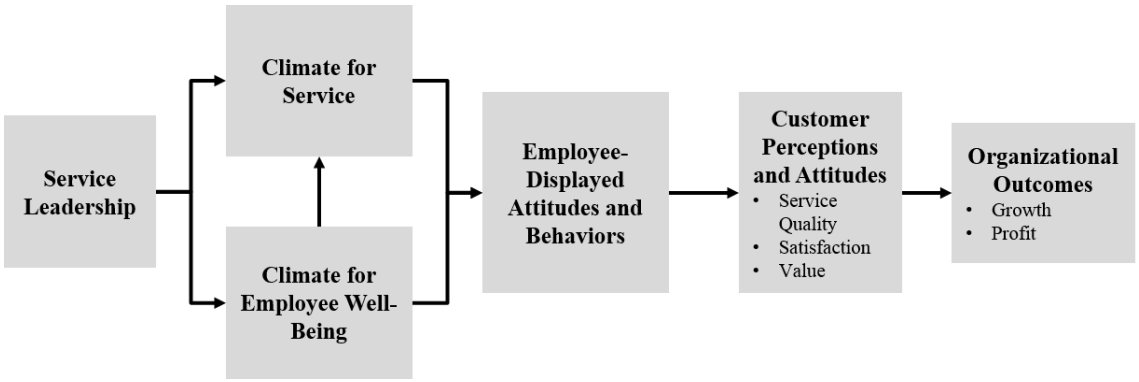


Figure 10. *Bowen's (2008) model of linkages in the Service profit chain (adapted from Bowen 2008, p. 164).*

Bowen (2008) presents the linkage model using six constructs. He suggests that the service organization's leadership creates the necessary climates in the service organization, i.e. the climate for employee well-being and the climate for service. The climate for employee well-being results in a good working environment and is the foundation for the climate for service. Leadership creates the climate for service to enable the employees to meet customer needs. These climates show in the perceptions of the employees about their organization and affect their attitudes and behaviors. Employee attitudes and behaviors affect the customers' perceptions of service quality. Finally, customer satisfaction leads to organizational outcomes such as growth and profit. (Bowen 2008, pp. 163-166.)

It should be noted, that the constructs in the linkage and SPC models can be formed in different ways. For example, Gelade et al. (2005) use commitment as a measure for employee (displayed) attitude, Homburg et al. (2009) use employee job satisfaction and Schneider et al. (2005) use customer-focused organizational citizenship behavior (OCB). Most of the authors study only one or a part of the links described in the SPC (e.g. Bettencourt & Brown 1997; Salanova et al. 2005; Schneider et al. 1998; Schneider et al. 2003), while some examine the model and its implied relationships as a whole (e.g. Gelade & Young 2005; Homburg et al. 2009; Schneider et al. 2005).

Developing the work of Bowen (2008), Merja Fischer further elaborated Bowen's model in her doctoral dissertation, linking concepts in the existing linkage and SPC research to the respective constructs in the model. This conceptualization is presented in Figure 11. The constructs in the model are separated into internal (service supplier) and external

(customer), based on whether the construct is observed internally at the service supplier or at the customer organization.

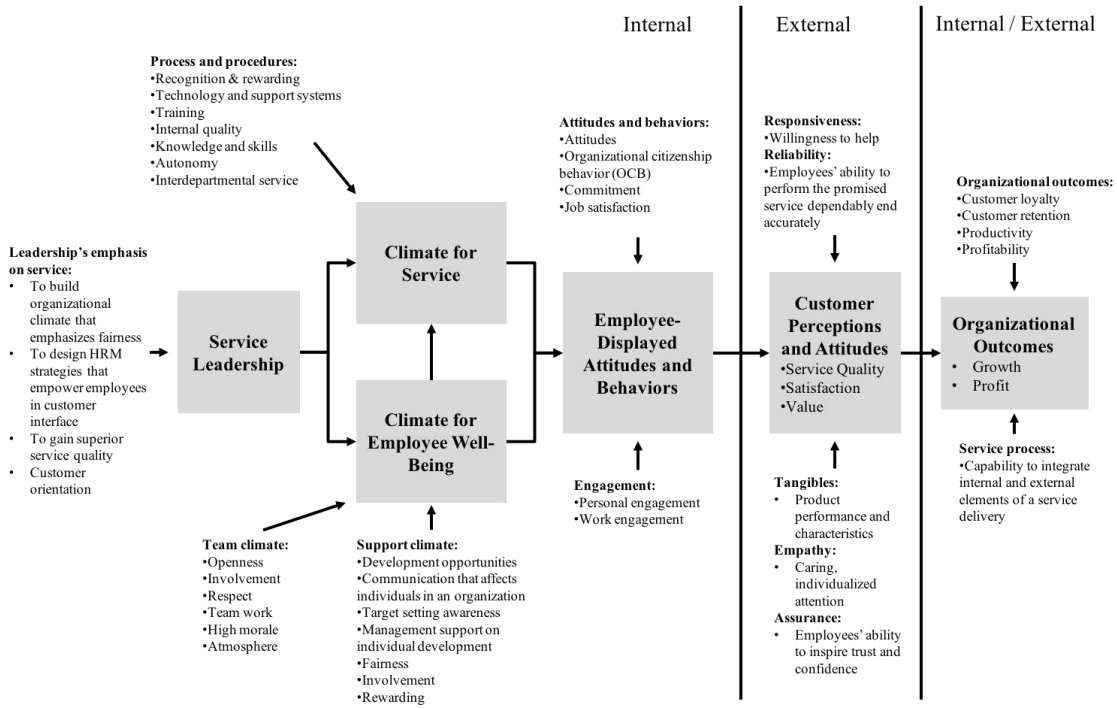


Figure 11. *Fischer's (2012) conceptualization of Bowen's (2008) linkage model (adapted from Fischer 2012, p. 42).*

In addition, Fischer tested the linkage model with empirical data from equipment maintenance service in business-to-business context. Fischer used existing employee and customer satisfaction survey data. It is notable, that several other studies have also used existing data in testing the linkages (e.g. Schneider et al. 1998; Gelade & Young 2005). Fischer found evidence for the existence of the linkages described in Figure 11. More specifically, she found that field service employees' perceptions of workplace climate and internal quality have an effect on their personal engagement. Furthermore, field service employees' personal engagement predicts customers' perceptions of service quality. Field service employees conduct the maintenance work in the customer's facilities, and therefore interact with the customer employees. Fischer examined the links also using account managers, who are usually in contact with the customer by phone or email. In account managers' case a relationship between employees' personal engagement and customers' perception of service quality was not found.

Since the purpose of this research is to develop an approach for common service quality measurement, the SPC and linkage research provide a framework to link the service quality to the operations and processes of the supplier. Therefore, we are especially interested in the supplier's side of the linkage research, i.e. the supplier's practices that may influence service quality. Appendix I. combines studies of linkage research and the used measures for supplier practices in these studies.

From Appendix I it can be seen, that the measures in the studies differ quite much from each other; the differences are even bigger in practice due to the differing definitions and measurement items. Therefore, no clear preferences can be obtained for the measures of supplier practices based on the review. One frequently used measure in the supplier practices is job satisfaction. However, apart from Schneider et al. (1980), none of the reviewed studies provide support for the link between supplier employees' job satisfaction and customers' satisfaction. In turn, Hallowell et al. (1996) found that internal service quality (defined as employee satisfaction with the service received from internal service providers) is related to job satisfaction and customer satisfaction. Explaining this result, Hallowell et al. (1996, p. 28) state that "While job satisfaction may not lead to customer satisfaction directly, service organizations rarely have satisfied customers without having satisfied employees".

Based on the existing linkage research literature, Pugh et al. (2002) identify eight dimensions, which they suggest drive service climate, service capability and quality service. These dimensions are: customer orientation and service quality emphasis, management support, hiring, training, rewards and recognition, teamwork, support systems, and customer feedback. However, all of these drivers are not equally important in all organizations (Pugh et al. 2002, p. 82). This is a very common theme in linkage and service profit chain research altogether, which can also be observed from Appendix I: as the studies test the certain linkages in the model, the exact same measures and especially measurement items are rarely used. This means that even if the constructs are conceptually similar, they still differ from each other in practice. Therefore, the results of existing SPC studies cannot be directly used to indicate the supplier practices that have an effect on service quality: the limitations due to the particular context, industry, service, measurement items and data gathering methods are apparent. This is well reflected in the contradictory results on the existence of the linkages. Therefore, as also Pugh et al. (2002, p. 82) note, the value of linkage research is in "identifying which practices matter most to the customers of a particular organization".

Overall, the literature on service profit chain and linkages suggest that the operations and processes of the service supplier can have an effect on the service quality and customer satisfaction (Anderson & Mittal 2000; Pugh et al. 2002; Fischer 2012). Therefore, including these supplier processes in the measurement of service quality is justified, especially, when the measurement system is intended to be used jointly by the buyer and the supplier of the service.

It is important to note that there are also conflicting results on the working of the service profit chain. For example, Schneider et al. (2003, p. 849) examined the relationship between employee attitudes and organizational performance. Their results suggest that some satisfaction attitudes lead to organizational financial and market performance and some do not, while some employee attitudes seem to be a result of the organizational performance. Also Schneider et al. (1998) found a reciprocal effect for organizational

climate and customer perceptions of service quality. Explaining their results, Schneider et al. (2003, p. 849) suggest that it is too simplistic to assume a relationship between the satisfaction attitudes of employees and organizational performance, because the relationship is complex. The conflicting research findings further elaborate the complexity of measuring the links in the service profit chain, but they do not warrant the dismissal of the principle.

3.3 Supplier performance measurement

Performance measurement has been widely researched, especially in the management accounting literature (Cousins et al. 2008, p. 241). Neely et al. (1995, p. 80) define performance measurement as "the process of quantifying the efficiency and effectiveness of action". In their definition, efficiency measures how economically the firm operates to provide a given level of customer satisfaction. Effectiveness on the other hand measures to which extent the customer requirements are met. (Neely et al. 1995, p. 80.)

Performance measurement enables the planning, controlling and directing of the activities of the organization (Cousins et al. 2008, p. 241-242). One of the most famous frameworks for performance measurement is the balanced scorecard by Kaplan & Norton (1992). However, the majority of the work in performance measurement has focused on intra-organizational performance rather than on inter-organizational relationships (Cousins et al. 2008, p. 242). In addition, even though the well-known measurement frameworks, such as the balanced scorecard (Norton & Kaplan 1992), emphasize also other perspectives in addition to the financial, performance measurement has traditionally had a financial focus (Gunasekaran et al. 2001, p. 86). However, this focus is shifting to include also the impact of costs on, for example, customer service, productivity and quality (Gunasekaran et al. 2001, p. 86). For performance measurement to be effective, the measurement goals must be linked to organizational goals, and there has to be a balance between financial and non-financial measures (Gunasekaran et al. 2004, p. 335).

Supplier performance evaluation and monitoring is an important responsibility of the purchasing function (Talluri & Sarkis 2002, p. 4257; Cousins et al. 2008, p. 242). The evaluation of suppliers is related to the last phases of the purchasing process presented in section 2.4. (see Figure 5). Typically, supplier performance evaluation includes the evaluation of quality, delivery, cost, and technical and managerial capabilities of the supplier (Hahn et al. 1990, p. 5). With performance measurement, the buyer signals and educates the supplier on the important dimensions of performance, and can also direct improvement activities (Cousins et al. 2008, p. 242). It is likely challenging to find suppliers that are instantly able to meet the buyer's requirements, (Noshad & Awasthi 2015, p. 466) and therefore supplier performance measurement is used as basis for supplier development (Hahn et al. 1990, p. 4).

This thesis focuses on the measurement of service quality of the supplier. The purpose of the supplier service quality evaluation is naturally to provide information on the service quality, but also to enable the better development of the service. This must be taken into account in the development of the supplier service quality measurement, so that the used measures provide useful information about the service and its quality. The target of successful supplier quality performance measurement is to facilitate the collaboration between the buyer and the suppliers (Noshad & Awasthi 2015, p. 471).

3.4 The initial framework for supplier service quality

Based on the literature on service quality and service profit chain and the first meetings with the case company, an initial version of the framework for supplier service quality was constructed. Because the quality attributes and dimensions seem to be at least to some extent dependent on the service, using an already existing model of service quality was not considered appropriate for the purposes of this research. The framework was decided to comprise of four dimensions: supplier practices, relationship, process quality and outcome quality.

Supplier practices refers to the inner practices and processes of the supplier that can be thought to have an impact on the supplier's service quality. The supplier practices dimension is based on the service profit chain research. Also, the case company clearly saw the importance of the supplier's practices on service quality. Relationship dimension is concerned with the quality of the relationship between the buyer and the supplier. Relationship with the supplier was deemed crucial by the case company, and the importance of the relationship is even further stressed by the utilization of a common measurement system. Furthermore, the benefits of close buyer-supplier relationships are well documented and widely acknowledged (Giannakis 2007, p. 401). Process quality dimension is related to how the service is performed, whereas outcome quality dimension is concerned with how the buyer perceives the actual outcome that results from the service. The process and outcome quality dimensions were included based on Grönroos' (1982) definition of service quality. The process and outcome quality are also widely supported in the literature, as can be seen from Appendix H.

In the initial framework for supplier service quality, the dimensions further consist of specific factors. Table 3. presents the initial factors chosen for each of the quality dimensions and the studies that can be linked to those factors. Only factors characterized by subjective perceptions of individuals with a basis in literature are included in Table X. to make it more clear. Both the subjective and objective factors of service quality are presented in the final version of the framework in Section 5.2. For the supplier practices dimension, suitable factors were gathered based on the linkage and SPC literature. Factors for the relationship dimension were gathered from the literature on service quality and buyer-seller relationships. For process and outcome quality dimensions, suitable factors were gathered from the reviewed service quality models.

It should be noted, that all studies use somewhat different names and definitions for the factors. There are however significant similarities between the factors of the presented studies. For example, for the first supplier practices factor in Table 3., job enablers, the naming used in this study has been adopted from Gelade & Young (2005), but Schneider et al. (1998) use the term "work facilitation" for a very similar factor. The structure and definition of the factors varies in most of the studies, and the exactly same definitions have rarely been used. Moreover, due to the dissimilar definition of the factors, a factor used in one study can actually correspond to two or more factors in another study. For this reason, similar items that have been used to define one factor in a study, can be found from several factors in other studies. This phenomenon can be found in all the chosen factors throughout the dimensions. The terms used for the factors in Table 3. have been chosen by the researcher.

Table 3. *Initial factors of the four dimensions in the developed supplier service quality framework.*

Initial factors and the related studies	Dimensions of supplier service quality			
	Supplier practices	**Relationship**	**Process quality**	**Outcome quality**
	Job enablers (Hallowell et al. 1996; Schneider et al. 1998; Schneider et al. 2003; Gelade & Young 2005; Salanova et al. 2005)	**Supplier's ability to develop the service** Not based on literature	**Attitude** (Haywood-Farmer 1988; Dabholkar et al 1996; Brady & Cronin 2001; Gounaris 2005; Ko & Pastore 2005, Caro & García 2007)	**Reliability** (Parasuraman et al. 1991, 1994; Dabholkar et al. 1996; Brady & Cronin 2001)
	Workplace climate (Schneider et al. 1998; Gelade & Young 2005; Salanova et al. 2005; Schneider et al. 2005)	**Service recovery** (Hart et al. 1990; Zemke & Bell 1990; Grönroos 2007)	**Behavior** (Haywood-Farmer 1988; Dabholkar et al 1996; Lytle et al. 1998; Brady & Cronin 2001; Gounaris 2005; Ko & Pastore 2005; Caro & García 2007)	**Valence** (Brady & Cronin 2001; Brady et al. 2006; Ko & Pastore 2005; Caro & García 2007, 2008)
	Feedback (Schneider et al. 1980; Hallowell et al. 1996; Schneider et al. 1998)	**Communication** (Anderson & Narus 1990; Grover et al. 1996; Ellram 1991; Korpela 2015)	**Expertise** (Brady & Cronin 2001, Ko & Pastore 2005; Caro & García 2007, 2008)	
	Personal engagement (Schneider et al. 1980; Schneider et al. 2003; Salanova et al. 2005)	**Trust** (Anderson & Narus 1990; Ellram 1991; Kumar et al. 1995a, 1995b; Grover et al. 1996; Doney & Cannon 1997; Homburg & Garbe 2001; Huntley 2006)	**Responsiveness** (Parasuraman et al. 1985, 1988, 1991, 1993; Cronin & Taylor 1992)	

For the supplier practices, four factors were chosen for the initial framework: job enablers, workplace climate, feedback and personal engagement. Job enablers factor describes the conditions that the supplier employees have to perform the job. If the grounds for performing the job are poor, the performance of the employee, and the resulting service quality, cannot be expected to be very good. Schneider & Bowen (1993, p. 46) found that work facilitation is significantly related to customer's perception of service quality. Furthermore, from the five HRM practices they studied, work facilitation provided the most consistent information about customer experiences (Schneider & Bowen 1993, p. 42).

Workplace climate, feedback and personal engagement are based on Fischer's (2012, pp. 84-92) similarly appointed factors in her conceptualization of Bowen's (2008) linkage model (see Figure 11.). In this context however, the workplace climate is defined very broadly: it can be considered to roughly corresponds with both the climate for employee well-being and the climate for service in Bowen's (2008) model. According to Bowen (2008, p. 165), both of these climates aim to create a good work environment which results in both satisfied employees and good service quality. Several concepts can be related to climate for employee well-being and climate for service (Fischer 2012, pp. 51-52). For example, Bowen et al. (1999, p. 19) argue that fair treatment of service employees "spills over" to customers, resulting in committed and satisfied customers. The feedback that the supplier employees get from their work and performance can be seen as important in delivering quality service. Based on the feedback the supplier employees can improve their working and the quality of the service they provide. A high engagement of the supplier employees on the other hand is thought to result in good performance with customers (Salanova et al. 2005, p. 1218), and to predict the customers' perceptions of service quality (Fischer 2012, pp. 112-113).

For the relationship dimension, four factors were initially chosen: supplier's ability to develop the service, service recovery, communication and trust. Supplier's ability to develop the service was chosen as a measure of relationship because the case company saw it important to be able to see the effort and results of the supplier in service development. Service recovery is an important part of the service delivery process because failures and variance in services are inevitable (Hart et al. 1990, p. 150) due to, for example, the human involvement and the simultaneity of production and consumption (Boshoff & Leong 1998, p. 24). The ultimate goal of service recovery is that the dissatisfied customer continues to do business with the company (Schweikhart et al. 1993, p. 5), i.e. maintaining the relationship. Zemke & Bell (1990, p. 42) state that "The true test of an organization's commitment to service quality isn't in the stylishness of the pledge it makes in its marketing literature; it's in the way the company responds when things go wrong for the customer".

The importance of communication and trust in order for the relationship to work efficiently is easy to understand. Both communication and trust have been identified as constructs in relationship quality (Jiang et al. 2016). In a case study by Ellram (1991), both mutual trust and communication were among the frequently cited success factors in international partnerships. It has also been suggested that a long-term interactive relationship is vital in the success of outsourcing (Grover et al. 1996, p. 106).

The process quality dimension was deemed to consist of attitude, behavior, expertise and responsiveness of the supplier employees; these factors are meant to measure the performance of the frontline employees of the supplier. It has been suggested that service encounters influence the perceptions of service quality (Bitner 1990, p. 79). The four chosen factors are well present in the service quality literature. Brady & Cronin's (2001)

interaction quality dimension consists of attitude, behavior and expertise in their conceptualization of service quality based on their qualitative research and literature review. Similarly, Caro & Carcía (2007) use these three factors as indicators of personal interaction dimension, in addition to problem solving. One of the three service quality elements of Haywood-Farmer (1998) is the behavior of the serving staff, including, inter alia, friendliness, politeness and attitude of the staff. Responsiveness was included based on the work of Parasuraman, Zeithaml and Berry. Of Grönroos' (2007, pp. 89-90) seven criteria of good perceived service quality, four can be linked to the chosen factors of process quality: professionalism and skills, attitudes and behavior, accessibility and flexibility, and reliability and trustworthiness.

Outcome quality was initially thought to include reliability and valence based on the literature. Reliability was adopted from Parasuraman et al. (1991) to measure the dependability and accuracy of the produced outcome of the service. Of the five Servqual dimensions, reliability has been found to be the most important in service quality evaluation (Parasuraman et al. 1988, p. 31). The importance of reliability (as "service delivery system failures") has been observed also by Bitner et al. (1994, p. 103). Several studies support the use of valence as an indicator of outcome quality (e.g. Brady & Cronin 2001; Brady et al. 2006; Caro & García 2007). According to Brady & Cronin (2001, p. 40), valence "captures attributes that control whether customers believe the service outcome is good or bad, regardless of their evaluation of any other aspect of the experience". Valence is used because perceptions of the service outcome can be influenced by factors that are outside the control of the service provider (Brady & Cronin 2001, p. 40). For example, gamblers may evaluate their experience with a casino based on whether they won or lost rather than considering the behavior of the staff or the tidiness of the casino (Brady et al. 2006, p. 83).

4. DEFINING SUPPLIER SERVICE QUALITY IN INDUSTRIAL SERVICES

4.1 Data collection and analysis

In this research data was collected using four different techniques: literature review, documentary analysis, semi-structured interviews, and questionnaires. The literature review was described in Chapters 2 and 3. The purpose of the literature was to provide the foundation for the research and the supplier service quality framework. The documentary analysis involved mostly going through the service level descriptions of cleaning service and reports from the quality rounds of cleaning service. Also the case company's internet-based system designed for collecting service requests and claims for real estate services was used. The documentary analysis helped to understand the current state of the quality measurement of cleaning service, and it provided insights about the aspects of cleaning service that are currently emphasized. Some of the information was used also in the evaluation of cleaning service quality (see Section 5.2.1).

The cleaning service examined and measured in this thesis covers all the cleaning that the supplier performs in the production units. This includes for example the cleaning of the control room, office spaces, conference rooms, break rooms, lobby and the restrooms. The employees of the case company perform the cleaning of the actual production spaces, where the production equipment is located. Therefore, this is not included in the analysis. The supplier of the cleaning service is a global provider of facility services.

The empirical part of this research consists of two phases. First, in this chapter the determination of supplier quality in industrial services is presented. The determination is based on the literature review, the semi-structured interviews and the confirmatory survey that was conducted as to confirm the results of the interviews. A framework for supplier service quality is developed as a result. The measurement of supplier service quality in cleaning service using a survey and its results are presented in Chapter 5.

4.1.1 Semi-structured interviews

For this research, a total of ten interviews were conducted: eight interviews from the case company's side and two from the supplier's side. The interviews were conducted between June and August 2016. The purpose of the interviews was to gain further understanding of purchasing, quality in services and quality measurement of services. All of the case company interviews were conducted before the supplier interviews. The language of the interviews was Finnish. The interview questions for the case company and the supplier are shown in Appendices A and B, respectively. The interview questions dealt with the

information needs of the companies concerning services and their quality. All interviewees were also asked to define either industrial or cleaning service quality, depending on the interviewee. This was done in order to examine the service quality also in a more general level. Even though the measurement data would be gathered from cleaning service, the wider viewpoint of industrial services aids the future development and use of the service quality measurement. Furthermore, it would be important to know if there were substantive differences between cleaning service and industrial service quality. In the interviews it was emphasized, whether there were some aspects of the service or service quality that were not currently measured, but what they thought that should be measured.

The initial supplier service quality framework (see Section 3.4) was used in the interviews as an overview of the dimensions and factors affecting service quality. The framework was introduced at the end of the interviews, after the interviewee had answered all the other questions. This was done so that the contents of the framework would not affect the answers of the interviewees. The dimensions and factors of the framework were gone through, and opinions and comments were asked about the framework. The supplier service quality framework would be updated based on the interviews. The development of the framework is discussed in Section 4.2. The first update was done after the case company interviews, so that in the supplier interviews it would be possible to better compare the views of both sides on service quality. The final version of the service quality framework was developed after the supplier interviews. The final version of the supplier service quality framework is shown in Appendix C. The length of the interviews varied between 26 and 90 minutes. Due to time constraints, one interview lasted only 11 minutes. The interview questions as well as the supplier service quality framework were always sent to the interviewees beforehand. All interviews were recorded and subsequently transcribed.

The interviewees from the case company's side were chosen in collaboration with the contact person of the case company. The criterion for the interviewees from the case company's side was that the interviewees were either dealing with purchasing, production or they had some connection to cleaning service and its effects. Three interviewees were chosen from the Rauma production unit in order to get an understanding of the practical side of the cleaning service on the production facilities. One of the case company representatives was an outside consultant that was responsible for real estate services for the case company. For the case company interviews two slightly different set of questions were used: One focused solely on the quality of cleaning services and the other one on both the quality of cleaning and industrial services. This was done in order to see whether the factors affecting service quality were different in industrial services and cleaning service. The pure cleaning service quality questions were presented to the three representatives from the Rauma production unit, while other case company representatives were asked about both the industrial and cleaning service quality.

The interviewees from the supplier's side were chosen with the contact person of the supplier company. Interviewees from the supplier's side had knowledge about cleaning service in industrial context and about the quality and performance measurement of cleaning service. Even though only three supplier representatives were interviewed, these interviews were of great importance to the research. The supplier is a large global service company and the service in question is part of their core business. Therefore, the supplier had a vast knowledge about the quality and the quality measurement of the service. The same questions were asked from all the supplier representatives. Table 4. presents more detailed information about the interviews and interviewees.

Table 4. *Detailed information about the interviews and interviewees.*

Company	Interviewee description	Duration (min)
Case company	Head of production	26
Case company	Head of purchasing	40
Case company	Head of main purchasing category	40
Case company	Purchasing director	90
Case company (an outside consultant)	Director of real estate services	25
Case company	Production manager	11
Case company	Service manager	44
Case company	Department manager	31
Supplier	Business unit director	60
Supplier	Business director Quality manager	59

The interviews in this research were conducted as semi-structured. In semi-structured interviews the researcher has a list of themes and questions, but depending on the interview and interviewee questions may be omitted or added. Also the order of the questions can vary depending on the answers of the interviewee. (Saunders et al. 2009, p. 320.) Semi-structured interviews were chosen, since the purpose was to familiarize the researcher with the research subject and the current practices in the case company. Structured interviews would not have worked, since it was vital to be able to ask additional questions to gain further knowledge on the subject. Unstructured (in-depth) interviews on the other hand could have worked. However, due to the number of specific areas that would have had to be covered in the interviews, semi-structured interviews were deemed most suitable.

The interviews were conducted one-on-one, with the exception of one interview where two interviewees (from the supplier's side) were present simultaneously. Compared to one-on-one interviews, group interviews may lead to a more productive discussion since

the interviewees can, in addition to answering the questions, evaluate and challenge each other's answers. On the other hand, in group interviews it is possible that some interviewees publicly agree with the views of others, even though they personally disagree. Group interviews also need to be managed more. (Saunders et al. 2009, pp. 345-346.) All interviews were conducted face-to-face.

4.1.2 The confirmatory survey

After the interviews of the case company representatives, a survey was conducted in order to further confirm the factors identified in the interviews, and to identify the most important factors that would then be used in the actual measurement of supplier service quality. The confirmatory survey would further confirm the results to the following research sub-question "*What are the factors of quality in industrial services and what is the connection between them?*"

The survey was conducted using the TUT Webropol survey tool and it was sent to all eight of the case company interviewees. A link to the survey was sent via email directly to the respondents. The survey was kept open about one and a half months because the timing coincided with the summer vacation season. A total of three reminders were sent during the time the survey was open. From the total of eight respondents, six completed surveys were received. The language of the survey was Finnish.

The confirmatory survey was structured according to the developed supplier service quality framework. The main distinction was that the survey included only the subjective measures of the supplier service quality framework. Only the subjective measures were included because in this way the survey results could be used to determine the factors that would be included in the actual supplier service quality measurement survey. It was also gathered from the interviews, that the objective quality measurement of cleaning service is highly advanced. Also for that reason the efforts were directed towards subjective service quality measurement, since in that area the researcher could contribute the most. This was also the area that was lacking in the case company's measurement of purchased service quality.

At each dimension of the supplier service quality framework, the respondent was asked to choose three factors that he or she considered most important in cleaning service, and then to place those factors in order of their relative importance (first being the most important and third being the third most important). The respondent was also given a possibility to add a completely new factor to any of the four dimensions. Two executives of the case company were chosen to answer the survey from the industrial services point of view.

4.2 The dimensions and factors of quality in industrial services

In this section the results of the interviews are presented in order to further develop the supplier service quality framework. The supplier service quality framework is a general framework that presents the relevant dimensions and factors affecting industrial service quality. The results of the confirmatory survey are presented at the end of this section.

The supplier service quality framework can be presented as a hierarchical model consisting of dimensions and factors. An overview of the final version of the framework is presented in Appendix C. In the first level are the dimensions of service quality. These dimensions are further divided into factors. The dimension and factor levels are thought to be generalizable as such to any purchased industrial service, i.e. outcome quality monitoring as a factor of outcome quality is important in all (industrial) services. However, the way the outcome quality is measured, may vary depending on the service in question. The factors then consist of individual items or measures, depending on the nature of the factor: subjective or objective. The framework should not be considered as an exhaustive description of service quality, but rather as a helpful tool in recognizing relevant aspects in service quality.

The first version of the framework for supplier service quality was introduced in Section 3.4 (see Table 3.). In the first version of the framework, service quality was seen to comprise of four main dimensions: supplier's practices, supplier-customer relationship, process quality and outcome quality. All of the dimensions remained in the final version, even though the dimensions were restructured and factors were modified. Supplier's practices dimension was renamed to supplier capability, since it became evident in the interviews that also other factors related to the supplier were seen as affecting service quality. Supplier's practices still remained as a part of the new supplier capability dimension.

Supplier-customer relationship was initially thought to comprise of supplier's ability to develop the service, service recovery, communication and trust. Supplier's ability to develop the service was ultimately moved to supplier practices. This was done, because it was deemed that the supplier's ability to develop the service could be measured by asking about it from the supplier employees, rather than from both companies. Service recovery was ultimately excluded from the framework. Even though it is suggested that service recovery influences customer's satisfaction with the service (Spreng et al. 1995, p. 18) and customer loyalty (DeWitt et al. 2008, p. 276), it was not included as a separate factor in the framework. Ultimately, the way service recovery is organized by the supplier is reflect in the evaluations of process and outcome quality. Furthermore, service recovery can be seen as a complex construct which would require a dedicated measurement instrument. For example, Boshoff (1999) has created a measure for service recovery called Recovsat. Therefore, supplier-customer relationship dimension was ultimately deemed to comprise of communication and trust.

In the process quality dimension attitude and behavior of the supplier employees were combined to one factor, the customer-employee interaction. This was deemed appropriate, since distinguishing attitude and behavior might not be useful in practice, and because practically both of these are evaluated based on interactions between the buyer and the supplier employees. In the first version of the framework, outcome quality was thought to comprise of reliability and valence based on the service quality literature. However, this structure was later discarded, and these two factors were combined to form the perceived outcome quality factor. Some of the items in reliability and valence were used also in other factors. In addition, outcome quality monitoring, achieved benefits and costs were added to the outcome quality dimension. The dimensions of the final version of the supplier service quality framework are presented in Figure 12.

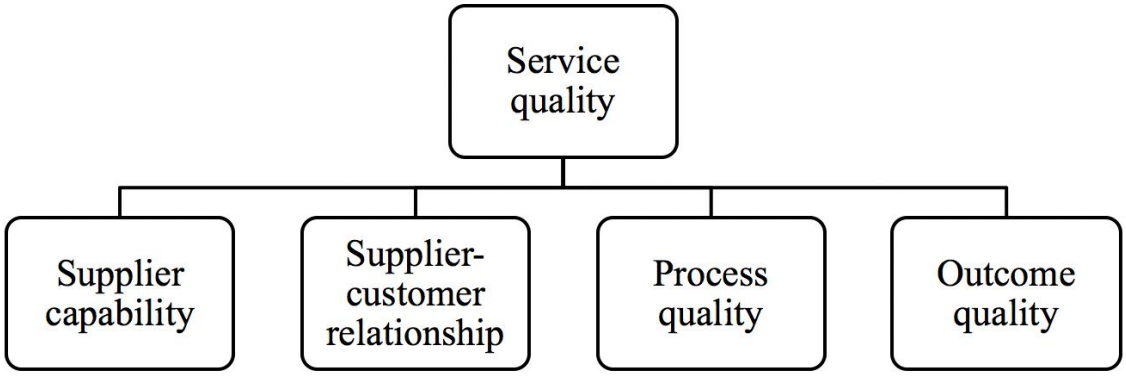


Figure 12. *The dimensions of supplier service quality.*

Supplier capability dimension was included, since the case company emphasized the importance of the supplier and its operations in the delivery of a service. This idea is based on the service profit chain theory in the literature. The supplier-customer relationship dimension was included, since a good relationship between the buyer and the supplier was seen as crucial by the case company representatives. The basis for this dimension was communication and trust between the parties. Process and outcome quality were included based on Grönroos' (1984) service quality model, and are meant to indicate how the service is perceived by the customer. The distinction between process and outcome quality is justified also based on the interviews: these were clearly seen as two different constructs. Process quality is concerned with how the service is delivered, i.e. the quality of the interactions between the supplier and customer employees, and the expertise and responsiveness of the supplier employees. The outcome quality dimension includes the factors that are related to the outcome of the service: outcome quality monitoring, perceived outcome quality, achieved benefits and cost-benefit analysis. The four dimensions and the respective aspect of service quality they represent, are shown in Figure X.

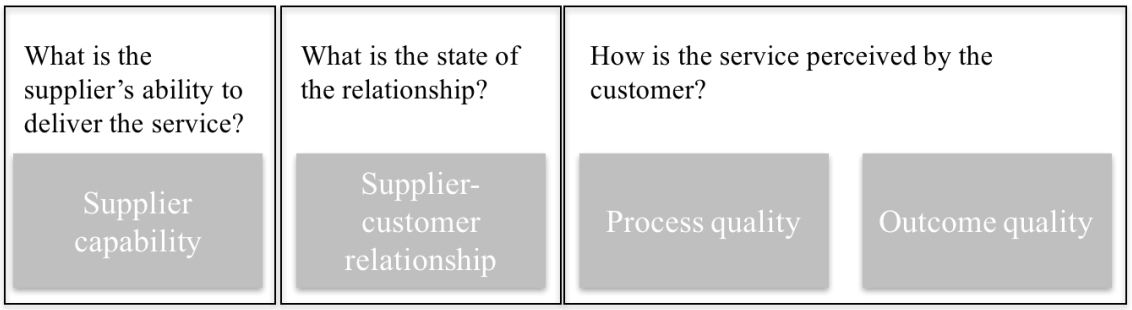


Figure 13. *The four dimensions of service quality and the quality aspects they represent.*

The dimensions in the framework were generally agreed to be important by all the interviewees. No clear preferences were present among the interviewees as to which of the four service quality dimensions is the single most important. However, when the interviewees were asked to define service quality before the introduction of the framework, outcome and technical quality of the service were mentioned most often, by seven of the eleven interviewees. The outcome and technical quality were mentioned by both the supplier and case company interviewees, and they were seen as an important part of service quality in both cleaning and industrial services.

The supplier capability dimension and the related factors are presented in Appendix D. The importance of supplier's practices and their effect on service quality was acknowledged by both organizations. The connection suggested by the service profit chain, that the service provider's operations and processes have an effect on service quality, was not new to the interviewees. Supplier's practices were eventually determined to comprise of job enablers, workplace climate, supervisor, training, supplier's communication, goal clarity, employee empowerment, personal engagement, and feedback. It should be noted, that this list is not exhaustive. These eight practices were chosen based on literature and in collaboration with the case company. It is possible that there are also other practices affecting the service quality (perceptions of customers). This is one important topic for future research. Furthermore, supplier's personnel indicators were included in the framework and they are seen as a result of the supplier's practices. Personnel indicators include employee satisfaction and employee turnover. About the supplier's practices' effect on service quality, the supplier's Business director said:

"[…] if you think about a logical chain, […] and to which our strategy is also based on, is that a committed and motivated personnel always correlates positively with good service quality and good customer satisfaction. And in our world that in turn has a strong correlation with a profitable customer relationship […]."

The case company saw that the actual execution ability of the supplier is of utmost importance. This of course has implications on the supplier's operations. Emphasizing the differences between suppliers, the Head of production of the case company stated:

"Ultimately, who is the best supplier, is about who can implement things to the operational level."

Organizational culture and company values of the supplier were seen as an important factor especially by the supplier representatives. Organizational culture and company values can be seen to affect supplier's operations and processes, as the Business director of the supplier stated:

"And company values. What are the company values and how do they show in everything? How do they show in recruitment, how do they show in employee orientation, training, feedback, development measures?"

A frequently mentioned factor was also the supplier's ability to develop the service. It was mentioned by both the case company and supplier interviewees. It was seen important that the supplier is able to continuously develop the service it provides. The case company interviewees considered the ability to continuously develop the service to be an indication of the importance of the delivered service to the supplier. The case company's Head of production said:

"Our principle is, that we would not want to buy a service from a company, if the service wasn't strategically important to that company, because then it (the service) will stay in its infancy."

The supplier-customer relationship dimension consists of two factors: communication and trust (see Appendix C). The importance of communication was acknowledged by most of the interviewees. About the relationship between the buyer and supplier, the Business director of the supplier stated:

"[…] These are the kind of things that it is important to decide these distinctly beforehand. For example, communication and information sharing […] require that the ways of working, meeting practices and information sharing practices are defined very clearly and in detail."

This implies that it is important to have well defined communication practices for the communication and relationship to work. If much of the communication is relied on informal information sharing, it may be that the communication is not adequate. Furthermore, at least in the case of purchasing services, the communication may not be effective, if the buyer and supplier have different kind of information available. A common measurement system facilitates the information sharing by giving both parties the same information, therefore making it easier to discuss about the service quality. Trust was considered a prerequisite for a working relationship, Head of purchasing of the case company stating:

"Actually, I have always said that there is no business if there is no trust, and in this kind of people-intensive job the trust should go all the way from the operative level to the senior management."

The process quality dimension in line with Grönroos' (1984) functional quality reflects the process in which the outcome is produced. The process quality dimension is presented in Figure 14. and it consists of customer-employee interaction, expertise and responsiveness. The quality of the interaction between the contact employees and the customer has a significant impact on service quality (Salanova et al. 2005, p. 1218). Like Ko & Pastore's (2005) client-employee interaction factor, the customer-employee interaction factor includes the supplier employees' attitude and behavior. In contrast to Ko & Pastore (2005), expertise is distinguished as a separate factor, since it was deemed important by many of the interviewees. For example, when asked about the most important things in cleaning service quality, the Department manager of the case company said:

"It is probably expertise. It is the expertise and the quality of cleaning."

The role of the supplier employees' attitudes in service quality were emphasized in the interviews. About the definition of service quality, the Head of production of the case company said:

"When thinking about cleaning service, isn't it pretty much so that the quality equals the employees' attitude towards their work?"

Based on the linkage and SPC research, the supplier employee's attitude should have an impact on the customers' perceptions of the service quality (see Figure X.), in this case specifically on the factors of process quality. Concerning service quality, the case company's Head of purchasing elaborated:

"And when we think about the supplier service quality, in these kind of people intensive things it is the professional pride, professional ability and professional want to do those things. […] And we all know, when talking about a carpenter or any other profession, that there are good carpenters and less good carpenters. […] And what the difference is, it is specifically the attitude towards the doing, and the attitude then correlates with the quality that is produced. And a professionally proud person produces good quality always regardless of the situation."

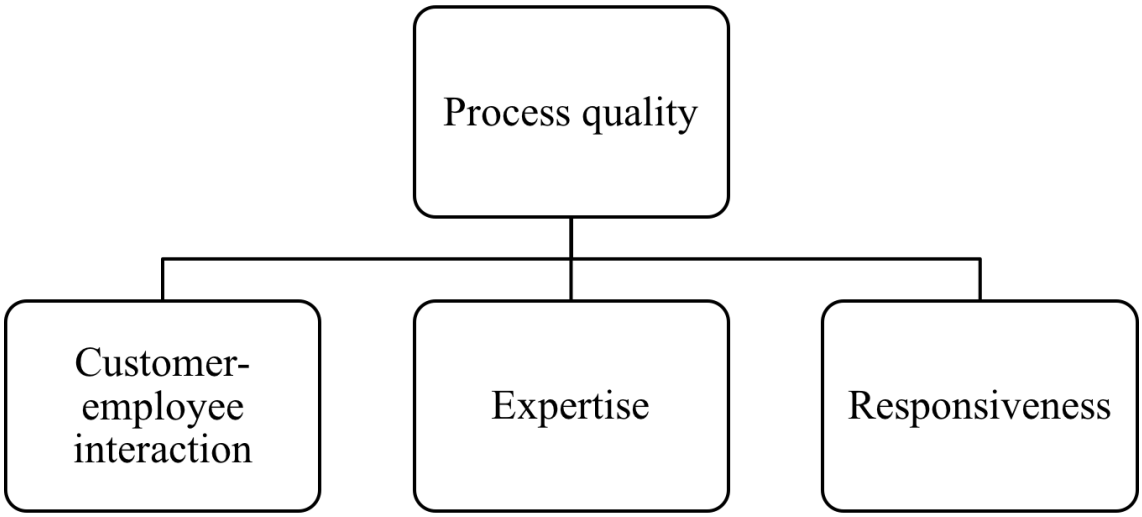


Figure 14. *The factors of the process quality dimension.*

A third factor in the process quality dimension is responsiveness of the supplier employees. Responsiveness is defined according to Parasuraman et al. (1985, p. 47) so that it "concerns the willingness or readiness of employees to provide service. It involves timeliness of service". The process quality dimension or one or more of the related factors were mentioned as a part of service quality by nine interviewees.

The outcome quality dimension includes the objective outcome quality monitoring, perceived outcome quality, achieved benefits and costs. The factors for outcome quality are presented in Figure 15. For almost all services there are objective measures (e.g. response time, frequency) for how the outcome of the service can be evaluated. Perceived outcome quality aims to capture the outcome of the service as it is perceived by the customer. Depending on the service, the relative importance of the objective and subjective outcome factors can vary. However, it is important to use both perspectives in quality measurement, so that a comprehensive picture can be formed. In the case of cleaning services, objective measures for the outcome quality monitoring include quality rounds, purity level assessments, cleaning frequency and the number of claims. Supplier's quality manager highlighted the difficulty of defining the outcome of cleaning service:

"Typically, when you think about it, […] that when the determination of the service quality takes a wrong turn, it is the use of the word "clean", […] because many different types of clean exist."

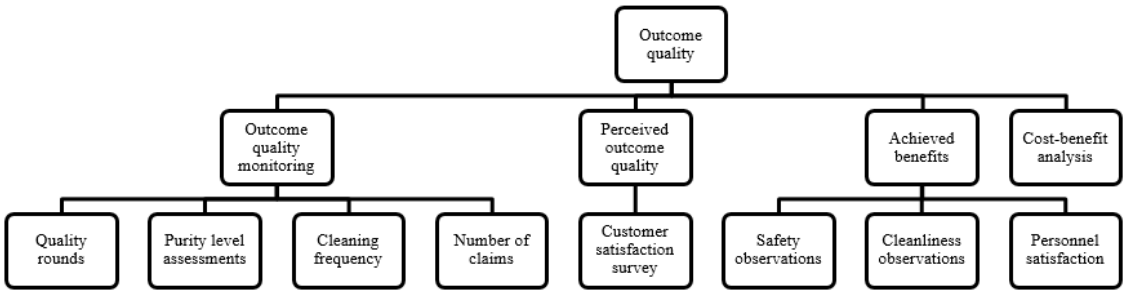


Figure 15. *The factors of the outcome quality dimension.*

It became evident in the interviews that currently cleaning frequency plays a vital role in the definition and contracting of cleaning service. The cleaning frequencies have been defined separately for each space in the production facilities, and the frequency is used to control the quality of cleaning. However, there is an incompatibility in using frequency in the management of the cleaning service, because in practice the frequencies are flexible. Moreover, another weakness of the cleaning frequency was reflected in the supplier's Business unit director's response:

"There is always the challenge, that when you have the frequency list and the (end) user has an assumption about something that is not related to the frequency […], so that what has been agreed and what has been bought […]."

There are clearly practical difficulties in using the cleaning frequency in the management of cleaning service. First, in some spaces the need for cleaning is not constant. This is due to, for example, the seasons and the varying use of the spaces. Therefore, following the predefined cleaning frequencies does not result in the intended quality: at a certain time, some of the spaces are cleaned too often while others are cleaned too infrequently. Second, the cleaning frequency does not adequately describe cleaning service quality, because there are also other aspects that have an impact on the quality of cleaning service that are not addressed by the frequency (for example, the equipment and the expertise of the supplier employees). Therefore, the cleaning frequency might not be a very useful measure of cleaning service quality, at least not on its own. Concerning the quality of cleaning service, the case company's Head of production stated:

"A lot of it is based on how the (buyer's) personnel that acts in those premises, how they experience the quality."

This view on quality is well in line with the literature's notion that the quality of a product or service is "*whatever the customer perceives it to be*" (Grönroos 2007, p. 73). Therefore, quality is ultimately defined and decided by the customer. Based on this, the buyer's employees' perceptions of the outcome of the service can be used as a measure of service outcome quality. Perceived outcome quality in cleaning service can be measured using a customer satisfaction survey for the actual end users: in this case, the case company

employees. The distinction between customer and end user is relevant, since also the satisfaction of the purchasing function could be measured. The satisfaction of the purchasing function would arguably be measured differently, and it would probably be more related to the relationship dimension. It is assumed that the continuous observations that the case company employees make about the tidiness of the facilities is reflected in the results of the customer satisfaction survey. Examining the satisfaction of the end users therefore offers useful information about the success of the service, and it can be considered a more useful indicator of cleaning service quality than, for example, cleaning frequency.

Achieved benefits includes the benefits that the service or the supplier provides to the customer, in addition to the technical outcome of the service. In the case of cleaning service, these achieved benefits could include safety and general cleanliness observations made by the supplier's employees, and also the more obvious customer's personnel satisfaction due to the clean facilities. The importance of these benefits became apparent in the interviews. For example, when asked to define the quality of industrial services, the external Director of real estate services of the case company said:

"[…] Generally the cleaners go to many places when they are working, […] what kind of observations they do outside their own sector […]. […] How do you observe, whether they are other tasks of your own company or some other company, how do you observe them and how do you forward them?"

The importance of achieved benefits was acknowledged from the supplier's side also. The supplier implied that the objective is to be more to customers than just a "performer of cleaning". The supplier's Business director stated:

"[…] How can cleaning participate, for example through observations of cleanliness, […] in addition to its own technical performing […]. […] So that we can be a part of an important strategic goal for Metsä."

The cost of the service was mentioned by practically all interviewees. It became evident that when considering (service) purchasing, a cost-benefit analysis is always a part of the analysis. Concerning the information needs related to industrial services, the Head of purchasing of the case company stated:

"[…] If I think about this from purchasing's point of view, the first thing is that do we get our money's worth from what we are paying for the service."

This might be due to the fact that one of the tasks of the purchasing function is to control costs (van Weele 2014, p. 53). Also, the purchasing company wants to make sure that it is not paying too much, and that might be why the interviewees saw costs and service quality as closely related. There seemed to be a consensus of opinion among the interviewees that service quality is good when the outcome and costs are at the level that

was agreed upon. The cost-benefit analysis of the service was therefore included in the outcome quality dimension, even though the economic aspect was present in only one of the reviewed service quality models (see Appendix H). When asked to define service quality of industrial services, the case company's Purchasing director clearly saw quality and costs going hand in hand:

"Generally, […] it is benefit vs. cost; that is the wanted quality achieved with that cost or are you ready to pay more."

Due to the fact that industrial services are generally produced and consumed in the customer's facilities, a distinct "physical environment" factor was not included in the framework, even though in the consumer markets this has been found to have an effect on service quality and customer satisfaction (Bitner 1990; Brady & Cronin 2001; Caro & Garcia 2007). However, it is important to note that the buyer's environment and facilities do have some effect, since the supplier's (and the customer's) personnel work in these facilities. Even though the customer's facilities may not have a straight connection to the supplier service quality, the facilities and environment may have an effect on a number of related things, such as the ability to complete the tasks and achieve service goals (Bitner 1992, p. 67). Therefore, it is important to ensure that the physical environment does not hamper or inhibit the delivery of the service.

All the service quality factors introduced in this section are originally based on either the existing literature, interviews or both. Table 5. presents the original source of the factors of supplier service quality framework.

Table 5. *The supplier service quality factors and their original sources.*

	Supplier capability	**Supplier-customer relationship**	**Process quality**	**Outcome quality**
Factor and the original source	**Organizational culture and values** Interviews & literature	**Communication** Interviews & literature	**Customer employee interaction** Interviews & literature	**Outcome quality monitoring** Interviews
	Supplier's practices Interviews & literature	**Trust** Interviews & literature	**Expertise** Interviews & literature	**Perceived outcome quality** Interviews & literature
	Personnel indicators Interviews & literature		**Responsiveness** Interviews & literature	**Achieved benefits** Interviews
	Supplier's ability to develop the service Interviews			**Cost-benefit analysis** Interviews

As it can be seen from Table 5., all the factors were mentioned or referred to at least to some extent in the interviews. The supplier service quality framework was largely constructed based on the views of the interviewees to ensure the relevance of the framework for the case company. However, most of the factors were also identified from the literature, which further validates the quality factors. Practically only the objective factors were based solely on the interviews: supplier's ability to develop the service, outcome quality monitoring, achieved benefits and cost-benefit analysis. Of these, the supplier's ability to develop the service can be measured also subjectively.

The confirmatory survey results helped to indicate the most important factors in each of the quality dimensions for the case company. As was mentioned in the previous section, two respondents were asked to answer from the perspective of industrial services, while others took the perspective of cleaning service. Six responses in total were obtained: two from the industrial services' point of view and four from the cleaning service point of view. The combined results are presented in Table 6. It should be noted that the framework used in the confirmatory survey was based on the initial version of the framework (see Table 3.), and therefore the structure does not fully correspond to the final structure of the supplier service quality framework.

Table 6. *The combined results of the confirmatory survey.*

Supplier practices	Supplier-customer relationship	Process quality	Outcome quality
1. Job enablers 2. Personal engagement 3. Feedback	1. Communication 2. Supplier's ability to develop the service 3. Trust	1. Attitude 2. Responsiveness 3. Expertise	1. Reliability 2. Valence

Job enablers was seen as the most important factor in the supplier practices in both cleaning and industrial service. Personal engagement of the supplier employees was seen as the second most important and feedback for the supplier employees as the third most important factor in cleaning service, while in industrial services the order was reversed. In the supplier-customer relationship dimension communication was seen as the most important factor in industrial services, while in cleaning service supplier's ability to develop the service was considered as most important, communication being second most important. Communication was more important in both cleaning service and industrial services. This might be due to the fact communication is probably a more practical concept of the two, and communication's effect on service quality is more easy to see than trust's. Considering process quality, the attitude of the supplier employees was seen as the most important factor in both cleaning and industrial services. Reliability was seen as the most important factor of outcome quality in both cleaning and industrial services. In industrial services also costs and supplier's knowledge of service levels were brought up in the open questions. The relative importance of the four dimensions was not explored in the survey.

5. SURVEY STUDY ON SUPPLIER SERVICE QUALITY MEASUREMENT

5.1 Data collection and analysis

This section describes the undergone measurement process for the supplier service quality measurement. The construction of the surveys is presented along with information on the administration of the surveys. Also the utilization and results of exploratory and confirmatory factor analyses for the case company survey data are presented in this section.

5.1.1 Data gathering with supplier service quality surveys

Surveys are usually used for exploratory and descriptive research. A benefit of conducting a survey is that it allows the collection of a large amount of data with relatively small costs. Surveys are usually conducted using a questionnaire. Using a survey, quantitative data can be collected and these data can be subsequently used to, for example, modelling relationships between variables. (Saunders et al. 2009, p. 144.) An important advantage of a survey is also standardization: since the questions are preset, all the respondents are asked the same questions (Schmidt & Hollensen 2006, p. 138). Also, a survey is the predominant tool used to measure service quality in the literature. Therefore, a survey strategy and a questionnaire was chosen for this part of the research. There are also several drawbacks in using a survey strategy. A serious effort should be made to pilot to the data collection instrument, as well as to ensure a good response rate and the representativeness of the sample. The data gathered using a survey strategy is also narrower than with some other, specifically qualitative research strategies. This is due to the fact that the number of questions in a questionnaire is limited by the practicality of the survey. It also needs to be ensured that the questionnaire actually gathers the data that is needed to answer the research questions. (Saunders et al. 2009, p. 144, 361.)

The developed supplier service quality framework described in Section 4.2 was used as a basis for the actual measurement of service quality. In this thesis, perception measures of service quality were used as they have been shown to be superior to disconfirmation (Dabholkar et al. 2000, p. 167). The supplier service quality surveys were constructed mainly from existing items. The existing items were searched from the literature. Also Merja Fischer's (2012) and Tuija Korpela's doctoral theses were used as a source for survey items. The items in Fischer's doctoral thesis are originally from the customer satisfaction survey of Wärtsilä. Preliminary themes (factors) for the search of items were chosen based on the interviews and the results from the confirmatory survey. The factors are categorized into the four supplier service quality dimensions defined in the

framework: supplier capability, supplier-customer relationship and process and outcome quality. The four different dimensions also mean that practically three separate surveys are required to measure all the dimensions: supplier capability survey for the supplier employees, supplier-customer relationship survey for both supplier and buyer respondents, and case company survey for the case company employees. The preliminary factors are presented in Table 7.

Table 7. *The preliminary factors for the supplier service quality survey.*

Supplier capability	**Supplier-customer relationship**	**Case company**	
		Process quality	**Outcome quality**
Job enablers	Communication	Attitude	Reliability
Workplace climate	Trust	Behavior	Valence
Supervisor		Expertise	
Personal engagement		Responsiveness	
Feedback			
Supplier's ability to develop the service			

The item search was conducted during three weeks. The literature was searched based on the preliminary factors and the four dimensions. Suitable items were collected and immediately categorized into the appropriate preliminary factors. Also the factor from the original source was written down. Fischer (2012) served as a good starting point to find articles with items concerning the supplier capabilities (e.g. Hallowell et al. 1996; Gelade & Young 2005; Schneider et al. 2003). Items for the communication and trust factors were found in the literature concerning buyer-seller relationships (e.g. Homburg & Garbe 1999; Korpela 2015). Items for the factors in process and outcome quality dimensions were mainly searched from the service quality literature (e.g. Brady & Cronin 2001; Parasuraman et al. 1991). When searching for suitable items, it was discovered that in many articles the actual items used in the study were not presented in the article. This of course guided the item selection towards those articles where the items (or some of the items) were presented. Ultimately, the survey items were based on 19 different sources. The final items (and their original sources) used in the supplier capability survey, the case company survey and the supplier-customer relationship survey are presented in Appendices E, F and G, respectively. Tables X.-X. also provide the coding for each of the items to facilitate the presentation of results. The codes are S for the supplier questionnaire items, C for case company items and R for relationship questionnaire items. Five of the total 60 items were specifically developed by the researcher.

After collecting the items, the list was gone through and the preliminary factors were further specified. In some cases, it was deemed that a factor name in the original source was actually better, and then that factor name and description was used. Also some new

factors were included, based on further categorization. At the same time, the most suitable items with respect to the factors were chosen from the list. When the categorization was complete, an additional search for items was conducted, since some of the original factors had been changed. It should be noted, that this survey development and the updating of the supplier service quality framework was done somewhat concurrently. Therefore, the changes in the categorization of the factors of the survey affected the supplier service quality framework, and vice versa. All the final factors are presented in Table 8. The factors marked with an asterisk (*) were eventually excluded from the survey. The exclusion of workplace climate, supervisor, training and supplier's communication was decided jointly with the case company. It was deemed that these items might not be of relevance at the moment. However, these factors could be added to the survey in the future, if seen necessary.

Table 8. *The final factors for the supplier service quality survey. Factors marked with an asterisk (*) were not included in the survey.*

Supplier capability	**Supplier-customer relationship**	**Case company**	
		Process quality	**Outcome quality**
Job enablers Workplace climate* Supervisor* Training* Supplier's communication* Goal clarity Employee empowerment Personal engagement Feedback Supplier's ability to develop the service	Communication Trust	Client-employee interaction Expertise Responsiveness	Perceived outcome quality

The dimension in the Table 8. also indicates the respondent for the respective items in the factors. The supplier capability dimension factors are answered by the supplier employees that actually do the job, i.e. the cleaning personnel. The process and outcome quality factors are answered by the case company personnel, i.e. employees, supervisors and managers, that work in the facilities that the supplier employees clean. In the case of the cleaning service, they are the end users. The items in the supplier-customer relationship factors are asked from both the supplier and the case company. This makes it possible to examine possible differences in communication and trust between the parties. However, the respondent group for these items is different from the rest of the items: the items in the relationship dimension deal with the practices and meetings between the supplier and

the case company. Therefore, these items need to be directed to persons that are familiar with the relationship between the two companies. From the case company the management level of each of the production units was deemed appropriate. From the supplier's side the site managers and service supervisors were chosen to answer the questions regarding the relationship of the companies. It should be noted, that the site managers and service supervisors are regional, not production unit specific. From now on in this thesis, the survey directed at the supplier employees will be referred to as "supplier survey" or "supplier capability survey". The survey directed at the case company employees, supervisors and management will be referred to as "case company survey". The survey concerning the relationship of the two companies will be referred to as "supplier-customer relationship survey" or "relationship survey".

A basic version of the supplier's customer satisfaction survey was used in the end of the item selection to see what kind of things the supplier asks its customers about the cleaning service it provides. Especially the items concerning the actual outcome of the service were of great interest. At this point it was found that almost all of the relevant aspects in the supplier's customer satisfaction survey were already included in the selected items. Ultimately, two items were included in the survey based on the supplier's customer satisfaction survey. These items were "The supplier employees take initiative" and "The appearance of the supplier employees is neat".

Most of the items had to be modified to fit the specific context of this research. For example, in the supplier survey items "my team" was changed to "my working community", because it was deemed to be more appropriate. Also, when referencing to the supplier employees who do the actual job, "supplier employees", were "supplier" was replaced with the actual name of the supplier, was used throughout the survey. Many of the items also had to be specified, since they were deemed to be too broad in the testing phase of the survey. Examples in parentheses were added to the items where appropriate. The wording of the items in a questionnaire is important, so that it can be ensured that the responses are valid (Saunders et al. 2009, p. 383). The surveys were conducted in Finnish, so the items had to be translated from English. The originally found items were first modified in English to better suit the purpose and context of this research. Then the items were translated into Finnish and further modified. For this thesis, the final items were once again translated into English. In all the items a five-point Likert scale was used, where the respondent was asked whether he or she agrees or disagrees with the statement. The options were strongly disagree (1), somewhat disagree (2), neither agree nor disagree (3), somewhat agree (4) and strongly agree (5). Also a "no answer" option was provided.

Among the existing items there were also a lot of items where the respondent was asked about the state of two separate concepts in the same item. An example of this is "the tasks of my job and the objectives set for me are achievable" (Fischer 2012). Tasks and objectives are two separate concepts, and it can be that, for example the tasks are achievable while the objectives are not. However, there is no way to know this based on

the answer, since the two concepts are included in the same item. The item was therefore divided into two separate items: “I can perform the tasks assigned to me” and “I can achieve the objectives set for me”. Another example of this was “In my working community feedback and ideas are obtained from customers”. This item was modified to “we obtain ideas from clients”. Therefore, when included in the survey, the items with two distinct concepts had to be either broken down to two items, or be modified so that the item contained only one concept.

The items of a questionnaire need to be checked within the context for which they were meant to ensure that the items are not misread or misunderstood (Saunders et al. 2009, p. 383). For this reason, the surveys were piloted with a few of the intended respondents from both companies. This is especially useful in order to identify errors that are apparent only to the target group (Schmidt & Hollensen 2006, pp. 157-158). The respondents for the pilot phase were chosen by the case company contact person. The pilot phase respondent group consisted of respondents from two Units (1 and 2) and two personnel groups (supervisor and employee). From the supplier’s side the management level respondents were used as a test group for both the relationship and the supplier survey. This was done because since the supplier employees didn’t have personal company emails, testing with the intended respondents would have been too time consuming considering the schedule of the research. Also, the researchers of Tampere University of Technology were used to pilot test the surveys. Based on the piloting, mostly minor changes concerning the phrasing and expression of the items were made.

The surveys were constructed using Questback’s Digium Enterprise platform. The case company uses this survey tool, so to ease the future use and development of the surveys this was deemed the most suitable platform. The surveys were administered through the Digium Enterprise platform by sending a link to the survey via email to the intended respondents. Because it was found out that the supplier employees do not have personal company emails, the links to the survey could not be sent straight to respondents. Rather, the contact person of the case company coordinated the administering of the survey through the management of each production unit. The management made sure, that the supplier employees working at the facilities had a chance to answer the survey using a computer.

The measurement of cleaning service quality was decided to be conducted in six production units of the case company. These production units are referred to as Unit 1, Unit 2, Unit 3, Unit 4, Unit 5 and Unit 6. The respondents from each of the production unit of the case company were decided by the management of each production unit. They were asked to provide a list of names by the contact person of the case company. The list was supposed to contain names from three different levels of the organization: the employee, supervisor and management level. For the employee level a minimum of 20 names was required from each production unit. For the supervisor and management levels a minimum of five names for each were required. These requirements were mostly met,

but the number of persons from the management level varied between one and three depending on the production unit. This is understandable, since it was required from the management level respondents that they were familiar with the relationship between the case company and the supplier. The case company survey was sent to a total of 165 persons. Of these, 11 were management, 30 were supervisors and 124 were employees. The intended respondents for the supplier capability survey included all the supplier employees (cleaners) working at each production unit. The supplier-customer relationship survey was sent to eight persons in the supplier's side (the site manager and the service supervisor of each region of the production units) and to the 11 management level persons from the case company. A total of 102 responses were received to the surveys. The response rate for the case company, supplier and relationship surveys were 47 %, 67 % and 84 %, respectively. Two reminders were sent to the respondents concerning the case company and supplier-customer relationship survey. The number of responses from each Unit are presented in Table 9.

Table 9. *The responses and response rates of the surveys at Unit level.*

Personnel group	**Unit 1**	**Unit 2**	**Unit 3**	**Unit 4**	**Unit 5**	**Unit 6**	**Total**
Case company employees	5	10	9	10	5	11	50
Case company supervisors	2	4	3	4	3	4	20
Case company management	1	1	1	1	2	2	8
Total case company responses	8	15	13	15	10	17	78
Response rate	30 %	56 %	50 %	52 %	37 %	59 %	47 %
Supplier employees	6	3	3	1	3	0	16
Supplier employees' response rate	100 %	75 %	75 %	50 %	100 %	0	67 %
Supplier management	2	2	2	2	-	-	8
Supplier management's response rate	100 %	100 %	100 %	100 %	-	-	100 %

Background questions for the surveys included unit, gender, age group, working experience in current positions and working experience in the company. For the supplier-customer relationship survey respondents, an additional background question was "I am familiar with the relationship with the partner", to which the same 5-point Likert scale as to the actual survey items was used. For the case company respondents, the unit and personnel group were set as background information in advance. Table 10. presents the number of responses to each of the surveys.

Table 10. *Number of responses at survey level.*

Company	Survey		
	Supplier capability survey	**Case company survey**	**Supplier-customer relationship survey**
Case company	-	78	8
Supplier	16	-	8

The analysis of the responses was done using IBM SPSS Statistics software and Excel. The next section describes the factor analyzes applied to the case company survey data. The results of the surveys are presented in Section 5.2.

5.1.2 Exploratory and confirmatory factor analysis for the case company data

An Exploratory factor analysis (EFA) was implemented for the case company survey data, to further examine the existing factor structure. Exploratory factor analysis explores the data in order to identify potential constructs, and it can be used in theory development (Hair et al. 2010, p. 707). EFA is widely used, especially in psychological research (Fabrigar et al. 1999, p. 272). Since the initial factor structure (see Table 8.) was constructed by the researcher and hence, did not fully correspond to any existing structure, it is justified to use EFA to better understand the structure of the data (Gerbing & Hamilton 1996, p. 63; Fabrigar et al. 1999, p. 274).

Before conducting the EFA, the survey data was screened for missing data, unengaged responses and outliers. 78 responses were received for the case company survey. Missing data in this case means that the respondent did not answer or chose the "no answer" option for an item. Based on the analysis, it was decided that if the respondent had more than 25 % of missing data in his or her answers, the response was deleted altogether. In practice this means that if the response had five or more missing values (of 18 total items), the response was deleted. A total of 4 responses were removed due to too many missing values. Unengaged responses were searched for by examining the standard deviation of individual responses. The threshold for standard deviation was set at 0.30. As a result, five responses were deleted, four of which had a standard deviation of 0. In addition, the remaining data was visually reviewed for certain patterns in responses (e.g. 1,2,3,4,5,1,2,3,4,5, etc.), but none were found. Outliers are "observations with a *unique combination of characteristics identifiable as distinctly different* from the other data" (Hair et al. 2010, p. 64). However, in the case of our case company survey, it cannot really be determined whether a response is an outlier or not, especially because the answers are based on the opinions of the respondents. In the survey data, outliers can only be examined using three variables: age group, experience with current position and experience with company. In this case, an outlier would be, for example, if the respondent

had more experience with current position than with the company, or if the respondent belonged to the age group of "under 20 years old" and had more than 10 years of experience with either the current position or the company. However, no outliers were detected in the data. As a result of the data screening process nine responses were deleted, so the sample size was reduced to 69.

Nonresponse bias in the data was tested using three response groups based on whether the respondent answered the survey after the initial invitation (first group), the first reminder (second group) or the second reminder (third group). This is a very common extrapolation method for testing for nonresponse bias, where the respondents who answer the questionnaire later are expected to be similar to nonrespondents (Armstrong & Overton 1977, p. 397). Using ANOVA, a statistically significant difference (at the 0.05 level) in the means of the respondent groups was found in items C6 "*The supplier employees take initiative*" and C18 "*The quality of the cleaning service of the supplier is so good, that I don't expect to find the same from other organizations*" between the first and the third group. However, it should be noted that the sample size for the first group was 40 and for the second group 17, while for the third group it was only 8, which naturally affects the results. The effect size of response group was further examined with Partial Eta Squared. For item C6 the Partial Eta Squared was 0.093, while for item C18 it was 0.194. This means that for items C6 and C18, 9.3 % and 19.4 % of the variance is explained by response group, respectively. The effect size of the response group in item C6 is not very large, and given the substantially smaller sample size, the nonresponse bias was not deemed substantial. In item C18 the effect size is quite significant, but given that this item was ultimately excluded from the analysis based on exploratory factor analysis (later in this section), it did not affect the results.

After the nine responses were deleted, the sample had missing data only 2.6 percent. The missing data was also well under 10 percent in each of the items, items C5 "*The supplier employees inform our working community about problems concerning the cleaning service*", C6 "*The supplier employees take initiative*", C12 "*The supplier employees react to occurring problems*" and C14 "*The supplier provides the cleaning service at the promised time*" having the most missing data with 5.8 percent. When analysing the data with Little's MCAR test, the results (Chi-Square = 243.218, DF = 275, Sig. = 0.917) indicate that the missing data is missing completely at random (MCAR). This means that several remedies can be used for the missing data, without introducing bias into the results (Hair et al. 2010, p. 62). Missing values were estimated using the expectation-maximization (EM) technique in SPSS. Imputing missing values in this case is justified, so that a sample size large enough can be obtained for further analysis. Using only the responses with complete data, the sample size would have been only 54.

The remaining sample size (N = 69) was still considered to be sufficient (though not very good) for factor analysis. Generally, a sample size of 50 is considered as the minimum for factor analysis (Hair et al. 2010, p. 102), even though the recommendations vary a lot

(e.g. de Winter et al. 2009, pp. 147-150). One much used rule is the subject (response) to item ratio of 10:1 (Osborne & Costello 2009, p. 137), which would have meant a sample size of at least 180 in this study. However, also smaller ratios (5:1 and even 2:1) have been used (Osborne & Costello 2009, p. 137). The subject to item ratio in the case company survey data was little below 4:1. The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) for the data was 0.872, suggesting that the data is suitable for factor analysis (Schmidt & Hollensen 2006, p. 302).

Exploratory factor analysis was conducted using Principal axis factoring and Promax rotation. Principal axis factoring method was chosen due to the non-normality of the data (Fabrigar et al. 1999, p. 277). Because the missing values in the data were imputed, listwise exclusion (complete case approach) could be used. The EFA was also conducted on the survey data with the missing values using pairwise exclusion (all-available approach). The criteria for pairwise exclusion were met: the extent of the missing data was acceptable (under 10 percent) and the missing data was random (Hair et al. 2010, p. 48). The EFA results were practically the same in both cases, which suggests that the results are reliable, and that the data screening and missing data imputation did not affect the results.

The cleaning service quality was assessed by using 18 items (C1-C18, see Table 13.). These items were initially thought to form four factors from the process and outcome quality dimensions: customer-employee interaction, expertise, responsiveness and perceived outcome quality. However, the exploratory factor analysis resulted in a three-factor model. These factors were named as responsiveness, expertise and perceived outcome quality, based on their content. The responsiveness factor still measures the responsiveness of the supplier employees, i.e. the willingness and readiness of the employees to provide the service. The expertise factor measures the perceived expertise of the supplier employees. Based on the four items of the updated expertise factor (C4 "*The supplier employees are friendly*", C7 "*The supplier employees are competent*", C9 "*The behavior of the supplier employees is good*" and C10 "*The appearance of the supplier employees is neat*"), expertise consists of the friendliness, competence, behaviour and appearance of the supplier employees. Perceived outcome quality consists of four items: three items (C15 "*Generally, the quality of cleaning is as good as I expect*", C16 "*Overall, I'm satisfied with the cleanliness of the working spaces*" and C17 "*I'm satisfied with the cleaning service of the supplier*") were also in the initial factor, but item C18 "*The quality of the cleaning service of the supplier is so good, that I don't expect to find the same from other organizations*" was replaced with item C14 "*The supplier provides the cleaning service at the promised time*". The customer-employee interaction factor was dropped as a result of the factor analysis: most of these items were moved into the Responsiveness factor.

Three items (C1 "*The supplier employees are always willing to help me*", C11 "*The supplier employees perform the cleaning service promptly*" and C18 "*The quality of the*

cleaning service of the supplier is so good, that I don't expect to find the same from other organizations") were removed from the model based on the factor analysis. Only loadings above 0.30 were taken into account, since this can be considered as a minimally acceptable value (Hair et al. 2010, p. 118). Item C1's loadings were 0.471 and 0.422, and item C11's largest loading was only 0.349 (also cross-loaded to another factor with a loading of 0.315). Also item C18 cross-loaded (0.564 and 0.301) on two factors. The reliability of the factors was examined using Cronbach's alpha. Cronbach's alpha supported the exclusion of the items C11 and C18 (though not C1), as the values were slightly greater for the factors without the items. Both items' highest loadings were to the perceived outcome quality factor. The deleted items and their factor loadings are presented in Table 11. The exclusion of these items does not greatly affect the measurement of cleaning service quality as a whole. Item C1 "*The supplier employees are always willing to help me*" might not be that suitable to cleaning service in the first place, as was suggested by one respondent already in the piloting phase of the survey:

"[…] To my understanding, the objective and purpose is not to ask and request stuff from the cleaners. The work should be planned and systematic so that the resources are allocated based on the intended purposes, not by "call voting" during the work."

Table 11. *Summary of the deleted items based on the initial factor analysis.*

	Deleted item		
	C1 "The supplier employees are always willing to help me"	C11 "The supplier employees perform the cleaning service promptly"	C18 "The quality of the cleaning service of the supplier is so good, that I don't expect to find the same from other organizations"
Factor loadings (factor)	0.471 (Responsiveness) 0.422 (Expertise)	0.315 (Responsiveness) 0.349 (Perceived outcome quality)	0.301 (Responsiveness) 0.564 (Perceived outcome quality)
Cronbach's alpha			
With the item	0.896	0.881	0.926
Without the item	0.885	0.887	0.932

Similarly, the suitability of item C11 "*The supplier employees perform the cleaning service promptly*" to the context of cleaning service can be questioned, because this might be difficult to assess altogether. Moreover, the case company personnel may not even be aware of the schedule, as one respondent from Unit 4 noted in the open questions. The use of item C18 "*The quality of the cleaning service of the supplier is so good that I don't expect to find the same from other organizations*" also has some difficulties. First, the responses on this item are largely dependent on the content of the service, i.e. what is bought from the supplier. If the service content is narrow, then the absolute quality of the

service cannot be very good to begin with. In this case, the dissatisfaction is not actually caused by poor service provided by the supplier, but rather by the narrow content of the service. Second, the case company personnel answering this item might not have any experience about other suppliers of the same service, and even if they did, comparing suppliers this way is inaccurate, since the content of the contract has likely been changed along with the change of the supplier. Hence, it was deemed that the inclusion of item C18 does not offer any additional value to the analysis. EFA was then implemented again without these three items. The factor loadings and Cronbach alpha values are presented in Table 12.

Table 12. *Factor loadings from the exploratory factor analysis of the case company data. Three items (C1, C11 and C18) were excluded from the analysis based on cross-loading. Factor loadings under 0.3 have been excluded from the table.*

	Factor (Cronbach's alpha)	**Loadings**		
		1	2	3
Item	**Responsiveness (0.885)**			
C13	The supplier employees react to Metsä Group's requests	0.873		
C2	The supplier employees make the effort to understand my needs	0.790		
C3	The supplier employees seek the best for the customer	0.720		
C12	The supplier employees react to occurring problems	0.674		
C5	The supplier employees inform our working community about problems concerning the cleaning service	0.664		
C6	The supplier employees take initiative	0.643		
C8	The supplier employees are interested in our working community's opinion about cleaning service	0.631		
	Perceived outcome quality (0.932)			
C16	Overall, I'm satisfied with the cleanliness of the working spaces		1.011	
C17	I'm satisfied with the cleaning service of the supplier		0.946	
C15	Generally, the quality of cleaning is as good as I expect		0.943	
C14	The supplier provides the cleaning service at the promised time		0.533	
	Expertise (0.887)			
C9	The behavior of the supplier employees is good			1.055
C10	The appearance of the supplier employees is neat			0.747
C4	The supplier employees are friendly			0.696
C7	The supplier employees are competent			0.482

All three factors had eigenvalues greater than 1, and they explained 67.0 % of the observed variance. All of the remaining items loaded quite well on their respective factors. Hair et al. (2010, p. 117) suggest that for a sample size of 60, the factor loadings above 0.70 are significant. For a sample size of 70, factors loadings above 0.65 are significant. From Table X. can be seen, that items C5 (0.664), C6 (0.643), C8 (0.631) and

C4 (0.638) have factor loadings below the suggested loading of 0.70 for 60 respondents, but very close to the required level for 70 respondents (0.65). Because these items did not load to any other factors, the items were included in further analysis. The factor loadings of items C14 (0.533) and C7 (0.482) are somewhat more below the required level of 0.65 for 70 respondents. However, based on the content and the relevance of these items to service quality, these items were also included in the analysis. Especially item C7 "The supplier employees are competent" is an important indicator of the perceived expertise of the supplier personnel. All factors also have more than three items, as "a factor with fewer than three items is generally weak and unstable" (Osborne & Costello 2009, p. 138). The Cronbach alpha values for the responsiveness, expertise and perceived outcome quality factors were 0.885, 0.887 and 0.932, respectively. These values indicate good internal consistency, as Hair et al. (2010, p. 125) suggest 0.70 as the lower limit for Cronbach's alpha. An overview of the changes made to the model based on the exploratory factor analysis is presented in Table 13.

Table 13. *Overview of the changes to the measurement scale based on the exploratory factor analysis.*

Initial structure	**Changes**	**Structure after the exploratory factor analysis**
Customer-employee interaction C1 The supplier employees are always willing to help me C2 The supplier employees make the effort to understand my needs C3 The supplier employees seek the best for the customer C4 The supplier employees are friendly C5 The supplier employees are interested in our working community's opinion about cleaning service C6 The supplier employees inform our working community about problems concerning the cleaning service **Expertise** C7 The supplier employees are competent C8 The supplier employees take initiative C9 The behavior of the supplier employees is good C10 The appearance of the supplier employees is neat **Responsiveness** C11 The supplier employees perform the cleaning service promptly C12 The supplier employees react to occurring problems C13 The supplier employees react to Metsä Group's requests C14 The supplier provides the cleaning service at the promised time **Perceived outcome quality** C15 Generally, the quality of cleaning is as good as I expect C16 Overall, I'm satisfied with the cleanliness of the working spaces C17 I'm satisfied with the cleaning service of the supplier C18 The quality of the cleaning service of the supplier is so good, that I don't expect to find the same from other organizations	- The four-factor model was changed to a three-factor model: customer-employee interaction factor was dropped. Most of the items in customer-employee interaction factor were included into the responsiveness factor. Item C4 was moved to the expertise factor. - Item C14 was moved from responsiveness to the perceived outcome quality factor. - Items C1, C11 and C18 were deleted from the scale due to cross-loading and practical difficulties in the suitability and usability of the items.	**Responsiveness** C13 The supplier employees react to Metsä Group's requests C2 The supplier employees make the effort to understand my needs C3 The supplier employees seek the best for the customer C12 The supplier employees react to occurring problems C6 The supplier employees inform our working community about problems concerning the cleaning service C8 The supplier employees take initiative C5 The supplier employees are interested in our working community's opinion about cleaning service **Perceived outcome quality** C16 Overall, I'm satisfied with the cleanliness of the working spaces C17 I'm satisfied with the cleaning service of the supplier C15 Generally, the quality of cleaning is as good as I expect C14 The supplier provides the cleaning service at the promised time **Expertise** C9 The behavior of the supplier employees is good C10 The appearance of the supplier employees is neat C4 The supplier employees are friendly C7 The supplier employees are competent

Confirmatory factor analysis (CFA) with maximum likelihood parameter estimation was implemented in order to further test the structure obtained by exploratory factor analysis. EFA explores the data for an underlying structure, whereas CFA tests how well a given structure actually represents the data. With CFA, the validity of the proposed measurement model can be tested and confirmed (Hair et al. 2010, p. 707). In CFA, the model's validity was assessed using standardized factor loadings, average variance extracted (AVE), and the following goodness-of-fit (GOF) measures: Chi-square, comparative fit index (CFI) and root mean square error of approximation (RMSEA). The suggested values for each of the measures are listed in Table 14.

Table 14. *The interpretation of standardized loadings, AVE and goodness-of-fit measures in CFA (Hair et al. 2010, pp. 669-709).*

	Measure				
	Standardized loadings	AVE	Chi-square	CFI	RMSEA
Preferred value or level	At least 0.5 or higher, ideally 0.7 and higher	0.5 or higher indicates adequate convergence	Should not be significant	Above 0.90 for a good model	A value below 0.05 or 0.08 can be considered good

As a result of the CFA, item C14 "*The supplier provides the cleaning service at the promised time*" was removed from the model. The item had a standardized loading of 0.67, which is only slightly below the preferred level of 0.70. However, the item had the second lowest loading in the model (item C6's "*The supplier employees inform our working community about problems concerning the cleaning service*" loading was 0.59), and was therefore a potential candidate for deletion. Both the CFI and RMSEA for the model were better without the item C14. CFI with the item C14 was 0.868, and without 0.890. RMSEA was 0.137 with item C14, and 0.131 without. The item C14 had also a relatively low loading in the EFA (see Table 12.). Furthermore, the applicability of the item in measuring cleaning service can be questioned partly on the same grounds as with item C18 "*The quality of the cleaning service of the supplier is so good, that I don't expect to find the same from other organizations*": the employees of the case company do not necessarily know the planned schedule of cleaning service. Even though item C6 had the weakest loading, the used measures did not unambiguously support its exclusion from the model: while the CFI of the model was slightly better without the item C6 (0.891 vs. 0.885), RMSEA of the model was worse without the item (0.146 vs. 0.141). Therefore, item C6 was retained in the model. The confirmatory factor analysis model and factor loadings after the deletion of item C14 are presented in Figure 16. using standardized estimates.

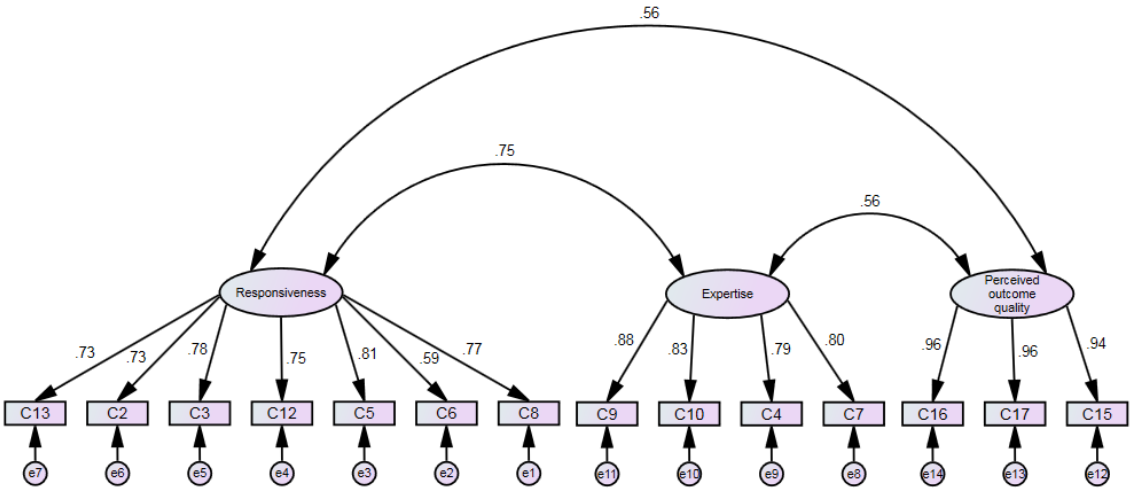


Figure 16. *The CFA model and factor loadings for the case company survey data (standardized estimates, N = 69).*

Overall, the factor loadings are very good. Only item C6 (0.59) had a loading under 0.7, but still higher than 0.5. Also the correlations between the factors were on an acceptable level, even though the correlation of 0.75 between responsiveness and expertise is quite high. This is however expected, since both factors measure the same higher order construct, process quality. The AVEs for responsiveness, expertise and perceived outcome quality factors were 0.55, 0.68 and 0.91, respectively, which suggest an adequate convergence (all are above 0.50). Overall, these results indicate a good convergent validity for the model.

Common method variance (CMV) was examined using Harman's single factor test (Podsakoff et al. 2003, p. 889). Common method variance is "variance that is attributable to the measurement method rather than to the constructs the measures represent" (Podsakoff et al. 2003, p. 879). CMV creates a false correlation among variables that is generated by their common source (Chang et al. 2010, p. 178). As all the data is from the same source, i.e. the case company survey, testing for common method variance is relevant.

Harman's single factor test is carried out in exploratory factor analysis to see if a single factor emerges or if majority of the covariance between the measures is accounted for by one factor (Chang et al. 2010, p. 180). 49.6 % percent of the observed variance was explained by one (responsiveness) factor (without item C14, since it was deleted from the model). Even though this can be considered high and possibly as an indicator that a substantial amount of common method variance is present, Podsakoff et al. (2003, p. 890) argue that there are no valid guidelines for the amount of variance that the first factor should extract. Also, the responsiveness factor consists of seven items, while the expertise and perceived outcome quality factors have four items. Arguably, this has an effect on the results of the Harman's single factor test. The inequality of the factors must be taken into account in the future development of the measurement scale.

Chang et al. (2010, p. 181) argue that the Harman's single factor test is insufficient to address the issue of CMV, and they recommend the use of multiple remedies. Common method variance was therefore examined using also common latent factor method, where the items are allowed to load on a latent common factor in addition to their theoretical constructs in confirmatory factor analysis. The significance of the structural parameters is then examined with and without the common factor to observe possible differences. (Podsakoff et al. 2003, p. 891.) The common latent factor method supported the finding that there is common method variance present in the data. This means, that the observed relationships between responsiveness, expertise and perceived outcome quality (i.e. process and outcome quality) are affected by the common data gathering method. Therefore, the results should be interpreted with caution. The observed CMV is especially problematic when the relationships between independent and dependent variables are examined using data from the same source. This should also be taken into account in the future development of the measurement. Ideally, common method variance is prevented by using multiple sources in the data collection, but also other methods for controlling CMV exist. (Chang et al. 2010, p. 178, 182; Podsakoff et al. 2003, p. 887.) The goodness-of-fit measures for the model are shown in Table 15.

Table 15. *Model fit measures for the CFA model.*

Measure		
Chi-square	CFI	RMSEA
159.7 with 74 degrees of freedom, significant at the 0.001 level	0.890	0.131

The Chi-square of the model is significant at the 0.001 level. CFI of the model is 0.890, which is only slightly below the preferred minimum level of 0.90 (Hair et al. 2010, p. 669). RMSEA for the model is 0.131. This is quite high, considering that the value should be at least below 0.08 (Hair et al. 2010, p. 667). Even though some of the measures for the model (Chi-square, RMSEA) are not at the desired level, the overall model fit can be interpreted as good. This indicates that the derived model presents the survey data adequately, and can therefore be used to measure the service quality of cleaning service. It is clear that the small sample size (N = 69) affects the results and limits their generalizability. The initial four-factor model (customer-employee interaction, expertise, responsiveness and perceived outcome quality) was also tested with CFA, but based on the used measures, the developed three-factor model was better. In addition, in the four-factor model the covariances between customer-employee interaction, expertise and responsiveness were all above 0.80, suggesting that these factors are very closely related. No other alternative model structures were developed for testing in CFA, because the EFA supported the chosen three-factor structure well.

For the supplier capability and supplier-customer relationship survey data factor analysis could not be implemented, due to the small sample sizes (16 respondents in both).

Furthermore, in the supplier data the number of items was actually larger than the number of responses (N = 16, 30 items), and the relationship survey had a very large amount of missing data (more in Section 5.2.3). A low subject to item ratio increases the chance, that the derived factors are sample-specific and therefore not generalizable. Moreover, with small sample sizes, findings should always be interpreted cautiously. (Hair et al. 2010, p. 102.) For the supplier and relationship surveys, the results had to be examined using the factors proposed according to the earlier literature (see Table 8.), not the empirical data gathered.

5.2 Results of the supplier service quality surveys

This section presents the results of the three surveys: case company, supplier practices and supplier-customer relationship. The results of each survey are described separately, along with discussion on the results.

5.2.1 Case company survey results

In all of the questionnaire items a 5-point Likert scale ranging from "strongly disagree" (1) to "strongly agree" (5) was used. The average time the respondents spent on answering the questionnaire was slightly under seven minutes. To examine the results of the case company survey, the mean values of the derived factors can be used. These are calculated from the mean of each of the items in that factor. Also, it is useful to look at the results at unit level, so that possible differences can be identified. The mean values for the responsiveness factor at unit level are presented in Figure 17.

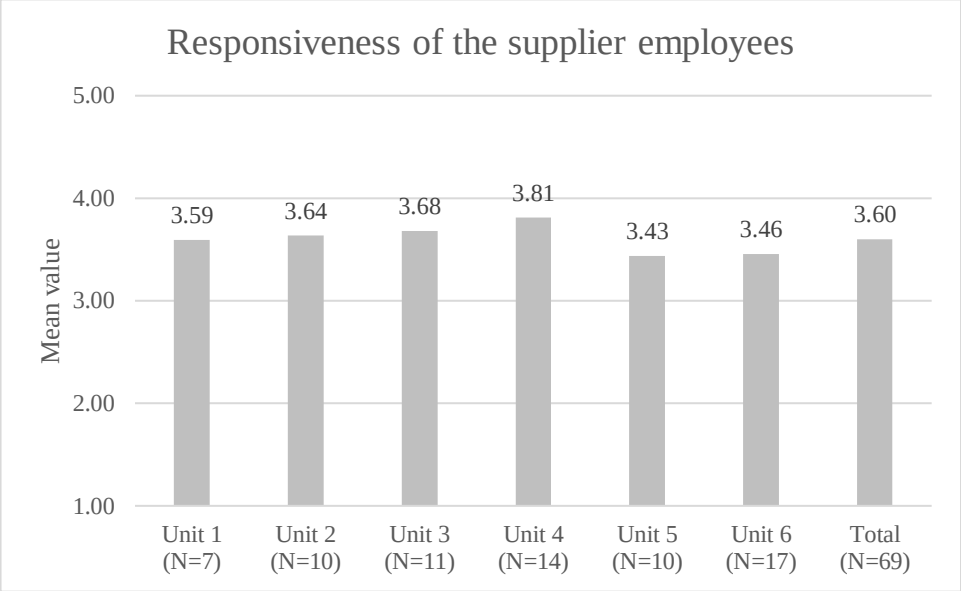


Figure 17. *Mean values for the responsiveness factor at each Unit.*

The responsiveness of the supplier employees seems to be at a good level in all Units. The highest score is at Unit 4, responsiveness of the supplier employees being 3.81. The

lowest averages are at Unit 5 and 6, the scores being 3.43 and 3.46, respectively. Still, even the lowest scores are clearly at the positive side (well over 3.00). The differences in the scores among the Units are not statistically significant. The average of all the Units is 3.60. Despite the good level of responsiveness in all of the Units, it became apparent from the comments to the open questions that there can still be substantial differences between the supplier employees of a certain Unit. One respondent from Unit 4 stated:

"The answers are based on an average from two supplier employees working at the production unit, it is difficult to answer (these questions) since one of them is good and the other one is not so good."

This is an important notion, since it is possible that a low score of a Unit is actually caused by the actions of only one supplier employee. It is also possible that the respondents answer the questions based on the "better" employee, which might cause the actual problems to stay hidden. This is clearly a restriction of the questionnaire as a measurement tool, and it emphasizes the importance of the comments provided in the open questions. The mean values for the expertise factor are presented in Figure 18.

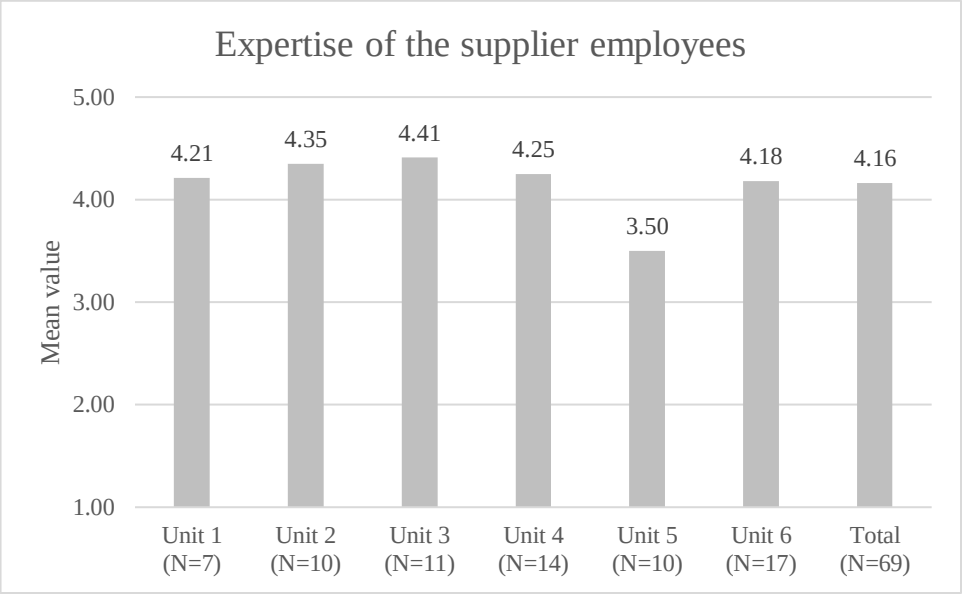


Figure 18. *Mean values for the expertise factor at each Unit.*

The expertise of the supplier employees is perceived as very good by the case company, the mean of all the Units being 4.16. However, the score in Unit 5 is 3.50, which is substantially lower than in other Units, even though still not statistically significant. In Unit 5, the expertise of the supplier employees is perceived as good (3.50), but not as good as in other Units. To an open question, one respondent from Unit 5 wrote:

"One of the supplier employees has a ragged shirt. This gives a bad overall impression, even though the person actually is active and competent."

Furthermore, another respondent from Unit 5 stated:

“One of the supplier employees is everything but customer-oriented. Doesn’t greet or talk, and appearance is serious. And the impression is that this person does everything with minimum effort. And for me, this person is “the face of the supplier”.”

The results then suggest that while the expertise of the supplier employees is perceived as very good in Units 1-4 and Unit 6, Unit 5 falls short in this regard. Based on this, some action should be taken to correct the situation. In this case, the comments to the open questions clearly point out what the issue might be (supplier employee’s appearance and behavior). When the service quality measurement is conducted again later, it can be seen whether the situation has been resolved, and if there are some other changes in any of the Units.

The comments also lend further support to the idea that process quality, i.e. how the outcome of the service is produced (Grönroos 1982), is an essential part of service quality. Furthermore, the comments suggest that the appearance and behaviour of the supplier employees has an effect not only on the perceived expertise of these employees, but also on the perception of the outcome they produce, as is reflected from the latter comment.

The answers to the open questions reflect well the observed difference between Unit 5 and other Units. For example, a respondent from Unit 3 (4.41) stated:

“I have no complaints about expertise, behaviour or appearance (of the supplier employees).”

The third factor measuring the cleaning service quality was perceived outcome quality. The mean values for the perceived outcome quality factor are presented in Figure 19.

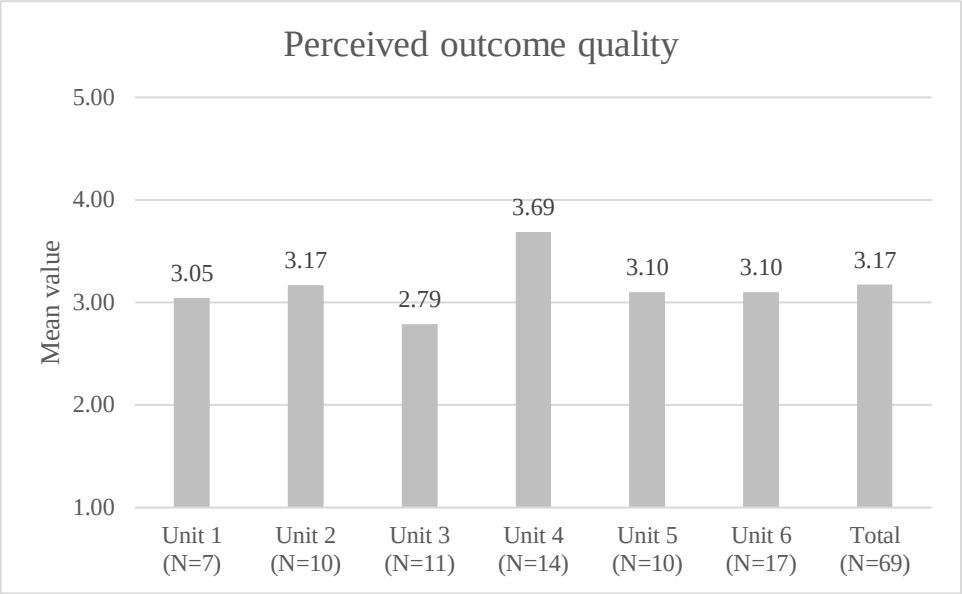


Figure 19. *Mean values for the perceived outcome quality factor at each Unit.*

The results regarding the perceived outcome quality factor are arguably the most interesting of the three factors. As already noted, the perceived outcome quality measures the actual outcome of the (cleaning) service as it is perceived by the case company (employees). Basically, this factor measures how satisfied the case company employees are with the quality of the cleaning service. As can be seen from Figure 19., Unit 3 has the lowest score (2.79) and Unit 4 has the highest score (3.69). The other Units are practically on the same level, a little above 3.00. Overall, the results on perceived outcome quality suggest that the case company's personnel are not satisfied with the quality of cleaning service. This is also reflected in the respondents' comments to the open questions. One respondent from Unit 3 noted:

"Ultimately the buyer decides the level of cleaning it wants. I hope they want better than this."

Based on the results on responsiveness, expertise and perceived outcome quality, the case company personnel are mostly satisfied with the supplier's employees. However, many of the respondents noted the impact of schedule to the cleaning service quality. The schedule of cleaning was perceived as simply too strict especially in Units 3, 5 and 6. Respondents from Unit 3 and 6 stated, respectively:

"The cleaners do a good job. The schedule is too tight, so the quality corresponds to this. Currently the cleaning is very superficial. This is not the fault of the cleaning staff, they do their best. […]"

"The schedule of cleaning has been designed to be so tight, that the cleaner does not have time for anything else but the necessary. In a dusty factory setting a more accurate and extensive cleaning would be more than welcomed for the sake of occupational wellbeing

and health. Supposedly, this is more the problem of the buyer or the service supplier, rather than of an individual cleaner."

Overall, it seems that the cleaners themselves are not the cause of the dissatisfaction on cleaning service quality. This supports the use of process and outcome quality as separate dimensions in the assessment of service quality: these clearly measure different aspects of the service, and the process and outcome of the service are seen as separate constructs also in practice by the respondents. This can be seen from the difference in scores: for example, in Unit 3, the outcome quality is perceived to be low (2.79), even though the expertise of the supplier employees is seen as very good (4.41).

Analysis of variance (ANOVA) was implemented for the data in order to see, whether the responses of different respondent groups had statistically significant differences. Between personnel groups, the managers of the case company gave on average the lowest ratings to the responsiveness (3.26) and expertise (3.66) factors, while employees gave the lowest rating in perceived outcome quality (3.24). The managers of the case company gave an average rating of 3.30 to the perceived outcome quality. The employees gave the highest ratings on the responsiveness (3.70) and expertise (4.23) factors. Supervisors gave the highest rating in perceived outcome quality (3.66). However, the differences between personnel groups were not statistically significant. Also, no significant differences were found between the responses of men and women or between different age groups of respondents. The unity of the responses in different Units can be examined with standard deviation. The standard deviations of responsiveness, expertise and perceived outcome quality in each Unit are presented in Table 16.

Table 16. *The standard deviations of responsiveness, expertise and perceived outcome quality in each Unit.*

	Unit 1 (N = 7)	**Unit 2** (N = 10)	**Unit 3** (N = 11)	**Unit 4** (N = 14)	**Unit 5** (N = 10)	**Unit 6** (N = 17)	**Total** (N = 69)
Responsiveness	0.74	0.88	0.89	0.68	0.88	0.61	0.75
Expertise	0.60	0.46	0.66	0.85	0.97	0.60	0.74
Perceived outcome quality	1.10	1.28	1.12	0.93	1.38	1.10	1.14

From Table 16. it can be seen, that the total standard deviations of responsiveness and expertise are approximately on the same level. The lowest standard deviation is in expertise in Unit 2 with a value of 0.46, which suggests that the respondents in Unit 2 agree quite well on the level of supplier employees' expertise in the Unit. The standard deviations of perceived outcome quality are consistently the highest, the total standard deviation being 1.14. This suggest that the respondents' views differ quite much in the assessment of perceived outcome quality. This is however expected, because the evaluation of the outcome of cleaning service is arguably affected by personal

characteristics and previous experiences of the respondent. Both the case company and the supplier interviewees acknowledged this. On the subjectivity of cleaning service quality, the Director of real estate services noted:

"Cleaning is an interesting service in the way that, for example, someone considers this room to be very messy and someone else does not."

Based on the information obtained from the case company representatives, the content of the cleaning service is approximately the same in all of the Units, and therefore also the costs are somewhat the same between the Units. Also the intended level of cleaning service is the same, even though there are some practical differences in performing the cleaning service between the production Units. This implies that in theory, the scores on perceived outcome quality should actually be on the same level in all of the Units. Of course, in practice this probably does not come true, due to the subjective nature of the measurement and the abstract nature of the measured construct. Everyone has somewhat different expectations and opinions about the cleaning service quality, and therefore two people might evaluate the same level of cleaning service differently. This applies to other services also. In addition, as long as it is not provided by a machine, the delivered service cannot be completely standardised. However, four of the six Units are practically on the same level, which lends further support for the use of perceived outcome quality as a measure of service quality. Only the highest score in Unit 4 is clearly on a different level than the others. This means that the case company employees are more satisfied with the cleaning service in Unit 4 than in other Units.

The inclusion of objective measures and aspects to the subjective service quality measurement might offer interesting additional insights. Even though the contents of the contracts are same in all of the Units, it does not tell much about what the actual level of the cleaning service is at each Unit. To examine the more objective quality of cleaning service, the results of quality rounds can be used. Furthermore, the perceived outcome quality can be compared to the corresponding quality round results, to find if there is any correlation between the two. The quality rounds are usually performed every or every other month. During the quality round, all predefined spaces are gone through in the production Unit, and all these spaces are graded on a scale from one to five, based on how well those spaces correspond to the predefined specifications. The scale is: "very much deviations" (1), "significant deviations" (2), "some deviations" (3), "good" (4) and "very good" (5). The scale is built so that if all the specifications are met, the quality is evaluated as very good. The quality round is performed by the supplier's service supervisor, but a case company representative is also allowed to participate. As a result, the quality round report lists the evaluation of all the spaces, and gives an average of the results. The perceived outcome quality and quality round results provided by the case company representatives and describing the results of October-November 2016 are presented in Figure 20.

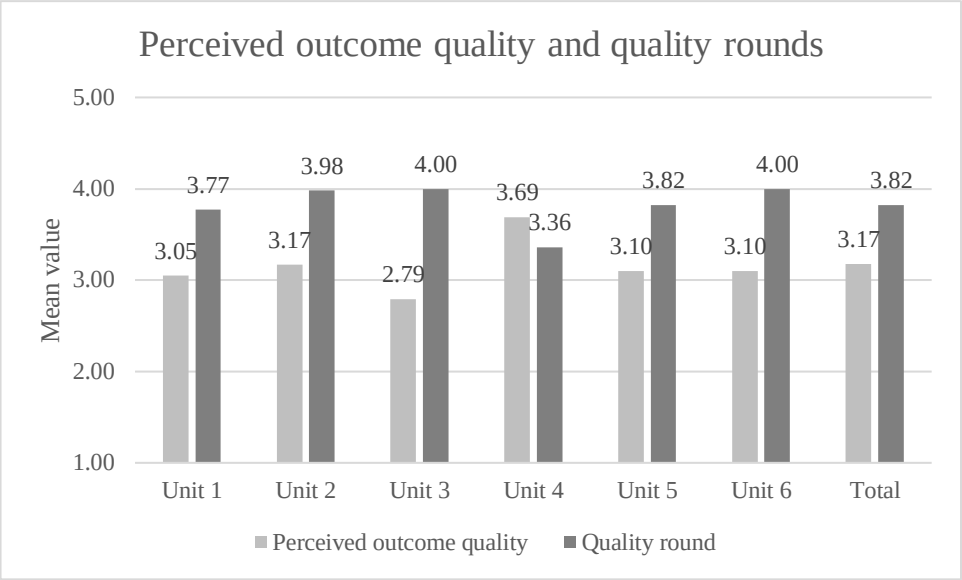


Figure 20. *Results of perceived outcome quality and quality rounds at each Unit.*

As the responses to the case company survey were collected mainly during October, the quality round results used for the comparison are either from October or November, depending on the Unit. As can be seen from Figure 20., the quality round results are systematically higher than the perceived outcome quality, except in Unit 4. Furthermore, Unit 4 has the lowest quality round result (3.36), while it has the highest perceived outcome quality (3.69). This is an interesting finding, because it seems that even though the cleaning service in Unit 4 does not fill the specifications, the case company employees are nevertheless quite satisfied with it. Overall, the quality round results seem to be good, the average being almost 4.00. However, this still means that all the specifications are not met.

As Figure 20. illustrates, there is no clear connection between the perceived outcome quality and quality round results. However, one cannot say much about the connection based on one measurement. To examine the possible connection between the quality round results and the satisfaction of the case company personnel, several measurement results are required from a longer period. Presumably, when the quality round results improve, the perceived outcome quality also improves. An interesting finding would also be that there is no connection, i.e. that an improvement in the quality round results does not result in an improvement in the perceived service quality. This would then suggest that the aspects assessed in the quality rounds have no effect on the end user satisfaction, which would somewhat question the purpose of the quality rounds.

Also the number of claims from each Unit could be linked to the score of perceived outcome quality. In this case, a claim is defined as a feedback from the case company which concerns tasks that are included in the contract, i.e. something was not done like it was supposed to. A high number of claims could be reflected as a low end user

satisfaction, and vice versa. All the claims are registered into a specific system, from where they can easily be accessed. However, the number of claims concerning cleaning service in the system seems to be extremely low. From October and November, only two claims were filed from all of the six Units in total. Therefore, no comparison could be done between the number of claims and perceived outcome quality. The number of claims is very low, which suggests that either there really are no claims, or that the claims are not input properly, or at all, into the system. The outside Director of real estate services noted in the interview, that the use of the system could be instructed better. Therefore, these practical challenges should first be addressed by the case company, before further analysis is possible.

If the contents of the contracts were different between the Units, then of course differences in perceived outcome quality could also be expected. In perceived outcome quality, the content of the service contract has to be taken into account when comparing the results of the Units. Perceived outcome quality differs from the other two factors in this regard. Responsiveness and expertise practically measure qualities of the supplier employees, and these should be somewhat on the same level regardless of the Unit in question (or the contents of contracts), especially since the supplier of the cleaning service is the same in all the Units. This also implies, that a direct comparison can be made between the Units in responsiveness and expertise factors. The results on responsiveness and expertise reflect quite well what was expected: when looking at Figures 17. and 18., the majority of the scores are very much on the same level. This suggests, that the supplier employees' responsiveness and expertise are quite coherent. In responsiveness, there are no distinct differences, while in expertise Unit 5 is on a lower level than other Units.

The perceived outcome quality could also be used as a tool in benchmarking: when the perceived outcome quality of a Unit is high (or on an otherwise desired level), then by comparing that Unit to another one with a (significantly) different score should reveal differences between the Units in some regard. Based on the comparison, action can be taken to get the outcome of the service towards the desired level. The inclusion of costs in this analysis could also reveal potential areas for improvement. The content of the service contract could be compared to the perceived service quality results, especially if data were available from a longer period. This might enable the case company to find the components of the service, which have the most effect on satisfaction and on perceived service quality.

5.2.2 Supplier capability survey results

Supplier capability was measured with six initial factors of supplier practices: job enablers, goal clarity, employee empowerment, employee engagement, feedback, and supplier's ability to develop the service. Due to the small overall sample size (N = 16), the results are presented using these six initial factors, since a factor analysis could not be conducted (see Appendix E). The results are also presented from all the Units combined.

The reliability of the initial scale was assessed with Cronbach's alpha. Table 17. presents the Cronbach alpha values for the six factors.

Table 17. *The Cronbach alpha values of the supplier capability factors.*

	Job enablers	Goal clarity	Employee empowerment	Employee engagement	Feedback	Supplier's ability to develop the service
Cronbach's alpha	0.728	0.805	0.649	0.815	0.912	0.938

The reliability of the initial scale appears to be decent, when looking at the Cronbach alpha values for the factors. Hair et al. (2010, p. 125) suggest 0.70 as the lower limit for the Cronbach alpha, and all factors except employee empowerment are above this. This indicates that the initial factor structure can be used to examine the data. The results of the supplier survey are presented in Figure 21. from all the Units combined. With all the responses combined, the results represent the supplier company as a whole. Missing data in the supplier questionnaire was 7.5 percent. One response was excluded from both the feedback and supplier's ability to develop the service factors due to large amount of missing data. Only a few of the respondents answered the open questions. Two respondents mentioned that they had difficulty answering the questions, since they had been working only for a couple of months.

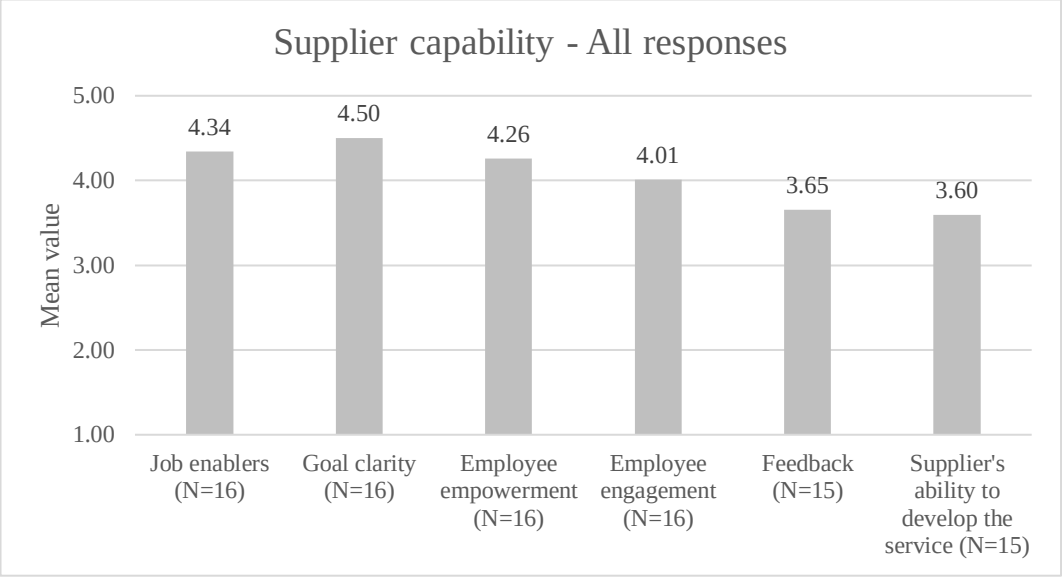


Figure 21. *The results of the supplier capability survey.*

From Figure 21. it can be seen that overall the supplier practices seem to be at a very good level. Job enablers, goal clarity and employee empowerment are above 4.00, which is very good on a five-point Likert scale. Only feedback and supplier's ability to develop the service are below 4.00. Considering that supplier's ability to develop the service was seen as important by the case company representatives, the mean value of 3.60 is probably

too low. Furthermore, if the supplier employees do not get sufficient feedback (3.65) from their performance, it might also indicate that the service is not developed and monitored adequately.

5.2.3 Supplier-customer relationship survey results

For the supplier-customer relationship survey, a total of 16 responses were received. Eight of these were from the supplier's side and eight from the case company. From the case company's side two responses were excluded from the analysis due to a large amount of missing data. The state of the relationship was assessed using two factors, communication and trust. Due to the low number of responses, a factor analysis could not be implemented to explore the underlying structure. Therefore, Cronbach's alpha was used to assess the reliability of the initial scale. The Cronbach alpha values for the communication and trust factors were 0.673 and 0.886, respectively. While the reliability of the trust scale seems decent, the Cronbach alpha value for the communication factor is somewhat below the preferred value of 0.70 (Hair et al. 2010, p. 125). This should be kept in mind when examining the results. The results on communication are presented in Figure 22. with all the responses combined.

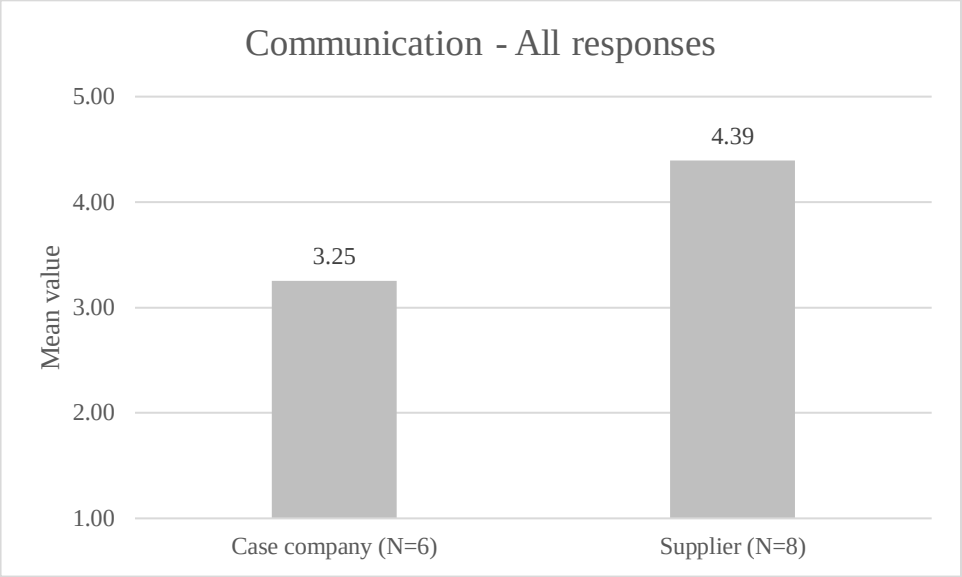


Figure 22. *The results of the communication factor of the supplier-customer relationship survey.*

As can be seen from Figure 22. the supplier's evaluation of the communication between the parties is on a much better level. The supplier responses have an average value of 4.39, which can be considered excellent. The case company responses have an average value of only 3.25, which is not good on a 5-point Likert scale. This suggests that the case company respondents are not satisfied with the level of communication between the companies. The results on trust are presented in Figure 23. with all the responses combined.

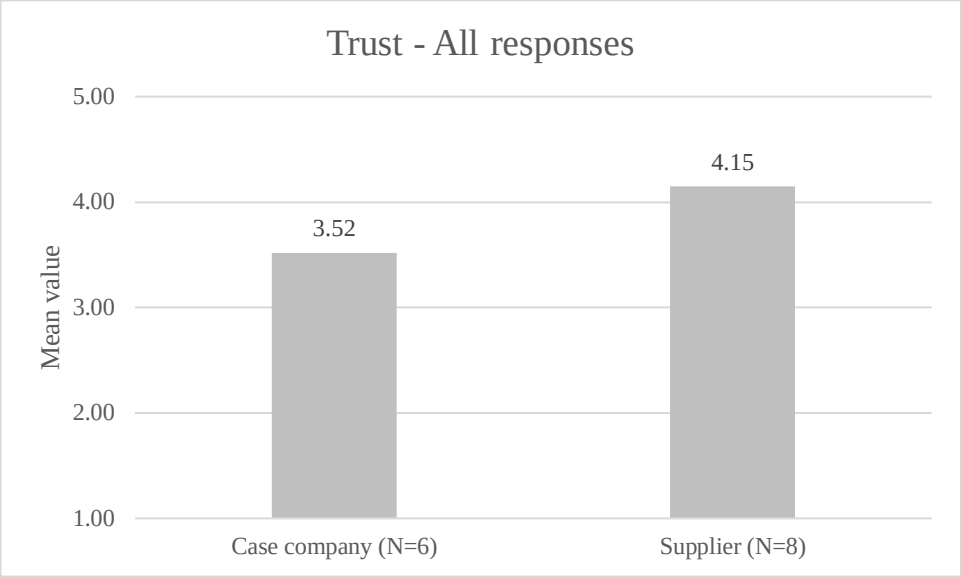


Figure 23. *The results of the trust factor of the supplier-customer relationship survey.*

As can be seen from Figure 23. the results on trust are quite similar to the results on communication. From the supplier's point of view, trust between the companies is good (4.15). The case company responses have an average value of 3.52. This suggest that the case company respondents think that there is not enough trust between the companies. However, the difference between the responses of the case company and supplier is clearly less than in communication.

The respondents were asked the following item as a background question: "I know well the state of the relationship between the case company and the supplier (e.g. contents of the contract, practices related to the cooperation)". The average of the six usable responses from the case company's side to this question was 4.7, which suggests that these respondents are familiar with the relationship. However, missing data accounted 21 percent of the total data in the six responses from the case company's side, while missing data in the supplier's responses was only one percent. This clearly suggests that the respondents chosen from the case company's side were not the right people to answer the questions regarding communication and trust between the companies. This finding is further supported by the answers to the open questions, as one respondent from the case company stated:

"The service manager of (Metsä) Fibre is actively and directly in contact with the supplier staff."

It seems that there are only a few (or possibly even one) people at each production unit, that are in direct contact with the supplier, and thus have knowledge about the communication and trust between the companies. These are the service managers of the

case company in each Unit. Therefore, considering the relationship, the service managers should be the respondents from the case company, while the site managers and service supervisors represent the supplier company. However, it is important to note that the site managers and service supervisors are regional, whereas the respondents from the case company represent a certain production unit. This means that if the case company has more than one production facility in a region, there is only a one set of corresponding responses from the supplier's side. This hinders the comparison of the responses.

5.3 A model of supplier service quality

Using the four supplier service quality dimensions developed in this thesis, service quality can be modeled as is presented in Figure 24. In the model, supplier capability (especially supplier practices) can be thought to have an impact on process and outcome quality. Furthermore, supplier capability might have an impact on outcome quality through process quality. The reasoning behind this is that if the supplier capability affects the process quality (i.e. responsiveness and expertise of the supplier employees), then this will likely also affect the outcome quality. For example, if the training of the supplier employees (supplier practice) is improved, this will likely results in a better expertise of the supplier employees (process quality), which in turn can have an effect on the perceived outcome quality (outcome quality). The supplier-customer relationship dimension can be seen as a moderator between the supplier capability and process quality dimension.

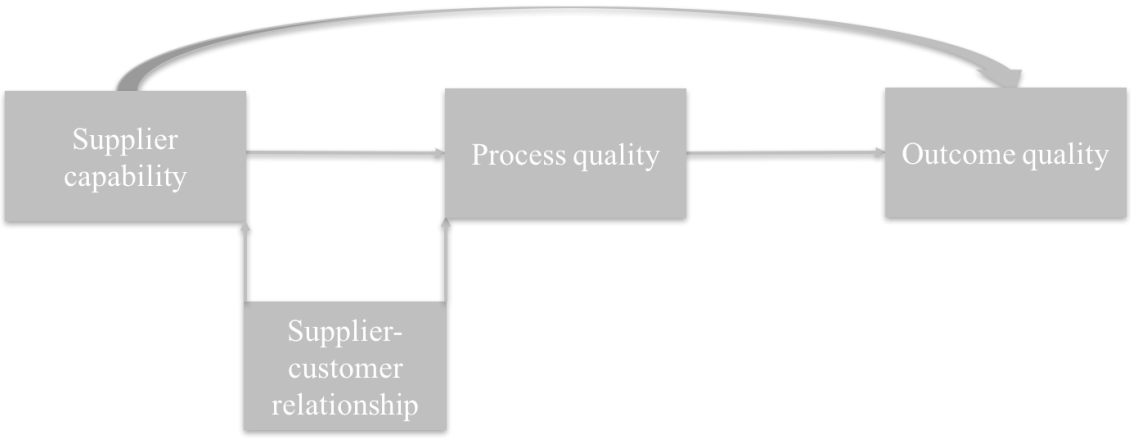


Figure 24. *The model of supplier service quality.*

However, due to the already discussed constraints of the supplier capability and the supplier-customer relationship questionnaire results, the hypothesized supplier service quality model could not be fully tested in practice. The model could only be examined regarding the process and outcome quality dimensions using the case company survey data. This part of the model is presented in Figure 25.

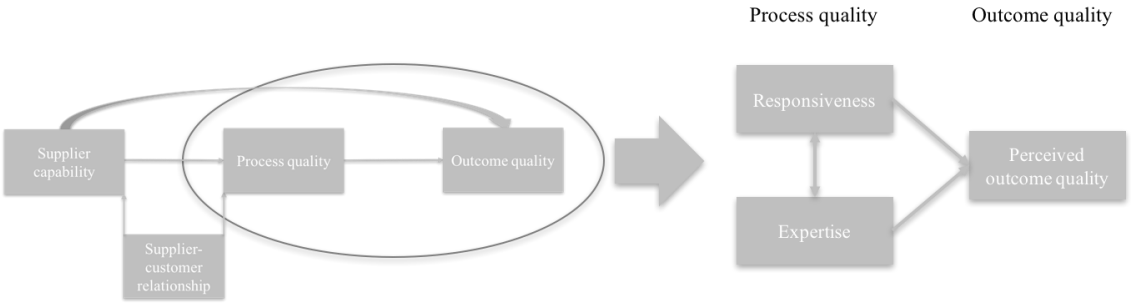


Figure 25. *The modeling of process and outcome quality dimensions.*

As presented on the right of Figure 25., responsiveness and expertise are seen as independent (i.e. endogenous) variables and perceived outcome quality is seen as a dependent (i.e. exogenous) variable. At first, the relationships between the factors of process and outcome quality dimensions were examined using Kendall's tau-b correlation. Kendall's tau along with Spearman's correlation is widely used in business and management research (Saunders et al. 2009, p. 461), and it was therefore deemed suitable for our analysis.

All of the 69 responses from the case company survey were used. The factor values for the correlation analysis were calculated for each respondent as mean of the respective factor items. The results of the correlation analysis are presented in Table 18.

Table 18. *Correlation analysis results of the factors of process and outcome quality using Kendall's tau-b correlation (N = 69).*

Factor	Responsiveness	Expertise	Perceived outcome quality
Responsiveness	-		
Expertise	0.543**	-	
Perceived outcome quality	0.411**	0.468**	-

** Significant at the 0.01 level (two-tailed)

Table 18. shows that all the correlations between the factors are positive and significant at the level $p < 0.01$. This provides support for the hypothesis that there is a connection between responsiveness, expertise and perceived outcome quality factors. Especially, the largest correlation is between responsiveness and expertise (0.543). This can be expected, since both of these factors measure the same higher order construct, process quality. It is also understandable in practice, that the case company employees do not assess the responsiveness and expertise of the supplier employees completely separately. Furthermore, the results suggest that there is a relationship between responsiveness and expertise and perceived outcome quality.

To further examine the relationship between responsiveness, expertise and perceived outcome quality, structural equation modeling (with latent variables) and path modeling

(with observed variables) were used. The hypothesized model is based on the depiction in Figure 25. The results of the structural equation modelling with maximum likelihood estimation are presented in Figure 26. The latent variables are depicted as ellipses.

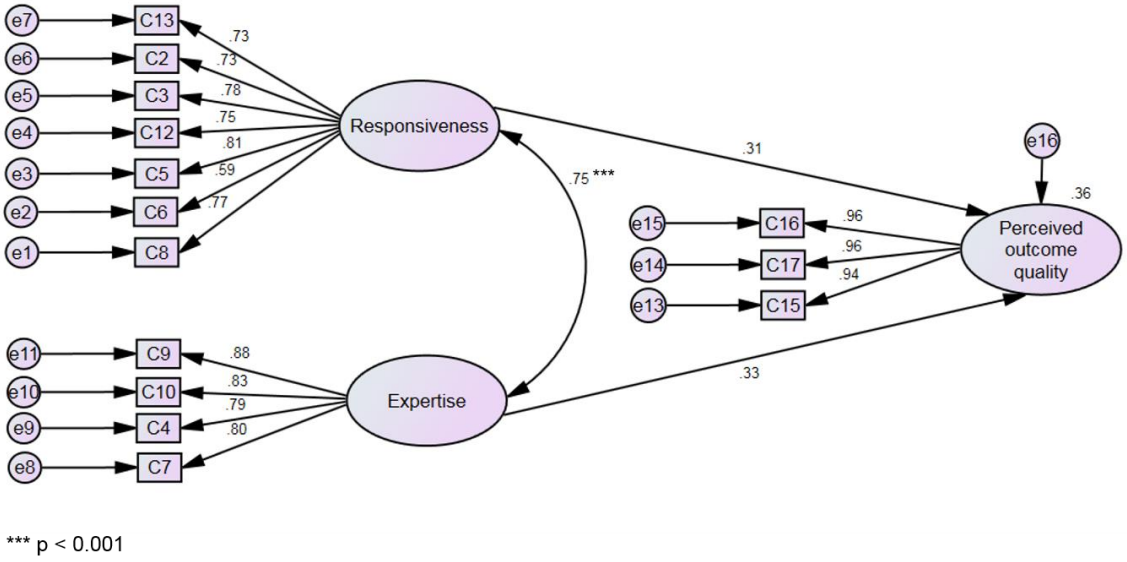


Figure 26. *The structural equation model and the respective results using latent variables (standardized estimates, N = 69).*

Due to the fact that the model in Figure 26. has the same number of structural relationships (three) as there are possible construct correlations in the CFA model (see Figure 16.), this model is considered a saturated structural model. For this reason, the factor loadings, the correlation between responsiveness and expertise, and also the model fit measures are equal to those of the CFA model. (Hair et al. 2010, p. 738.) The Chi-square is 159.7 with 74 degrees of freedom (significant at the 0.001 level), CFI is 0.890 and RMSEA is 0.131. Therefore, the overall model fit can be interpreted as good, as in the case of the confirmatory factor analysis.

As standardized estimates are used, the variable between responsiveness and expertise in Figure 26. (as well as in Figure 27.) is correlation. The standardized regression coefficients between responsiveness and perceived outcome quality, and between expertise and perceived outcome quality are 0.31 and 0.33, respectively. However, neither of these is statistically significant. The squared multiple correlation coefficient (R^2) for perceived outcome quality is 0.36, suggesting that approximately 36 % of the variance of perceived outcome quality is explained by responsiveness and expertise. This lends further support for the hypothesis that process quality has an effect on outcome quality. It also indicates that the perceived outcome quality is largely based on other attributes than responsiveness and expertise, which is expected.

Further examination is conducted with path analysis. The results of the path modelling with maximum likelihood estimation are presented in Figure 27. The path modeling is

conducted with observed variables, i.e. the values of the factors were calculated for each response as a mean of the respective factor items. The observed variables are depicted as rectangles.

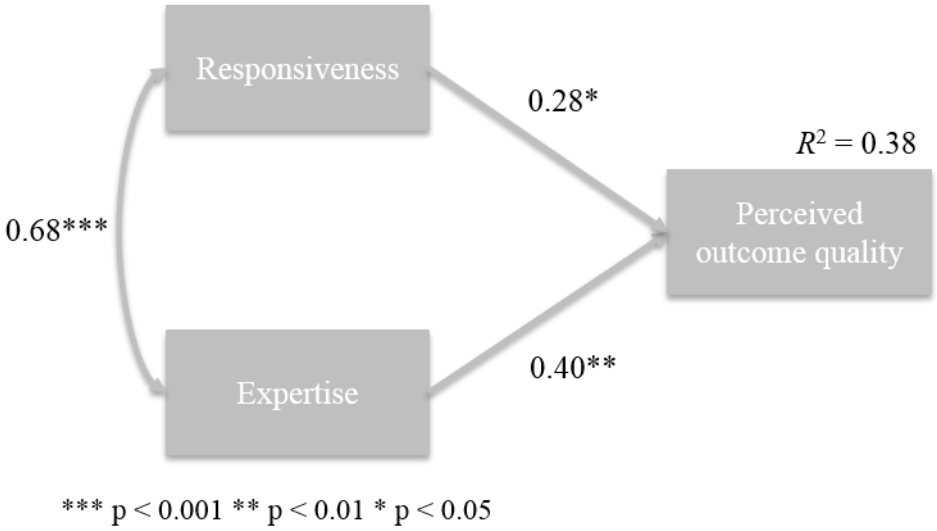


Figure 27. *The results of the path modeling using observed variables (standardized estimates, N = 69).*

The correlation between responsiveness and expertise is 0.68 and statistically significant, as it was in the structural equation model. The standardized path coefficient between responsiveness and perceived outcome quality is 0.28 and statistically significant at the level $p < 0.05$. Between expertise and perceived outcome quality, the standardized path coefficient is 0.40 and significant at the level $p < 0.01$. These results suggest that there is a stronger relationship between expertise and perceived outcome quality than between responsiveness and perceived outcome quality. This finding is expected, because it is understandable that the expertise of the supplier employees would have a greater effect on the perceived outcome quality than responsiveness. This finding implies that the outcome quality could be improved by improving the expertise (i.e. training) of the supplier employees. The squared multiple correlation coefficient (R^2) for perceived outcome quality is 0.38, which is close to the value obtained in the structural equation model (0.36).

In summary, structural equation modeling and path modeling results suggest that there is a relationship between responsiveness and expertise and perceived outcome quality, i.e. between process and outcome quality. Therefore, the results provide support for the examined part of the model. However, when interpreting the results, the results of the common method variance test (see section 5.1.2.) should be kept in mind. The results also indicate that the case company survey developed in this thesis can be used in the modelling and measurement of process and outcome quality dimensions.

5.4 Interpretation and the use of the supplier service quality measurement in the future

From purchasing function's point of view, consideration should be given to what the desired levels of the measurement results of a purchased service actually should be. This is largely dependent on the service in question. The classification of services that the buying company uses could be linked to the service quality measurement. The responsiveness and expertise of the supplier employees should probably be required to be quite high regardless of the service. Of course, if the service is only a necessity and low costs are the main objective, then a low score on responsiveness and/or expertise could be justified. The perceived outcome quality on the other hand is more dependent on the service in question. In some less important services, even a low perceived outcome quality can be tolerated; the reasoning mostly being that it costs less. The buying company should avoid paying for "extra quality", when a lower level of quality would be sufficient. In this regard, the measurement of end user satisfaction is however somewhat problematic. If in some service a mean value of 2.00 is accepted (or pursued), it means that the end users are not satisfied with the outcome of the service (compared to what they expect). The dissatisfaction of the end users is rarely desirable. Based on this, the score of the perceived outcome quality should never be under 3.00: at this point, the end users are not especially happy but not dissatisfied either. This could be considered the lowest acceptable level for any service.

However, the problems in interpreting the satisfaction of the end users stem from the fact that the expectations of the end user (company employee) and the customer (purchasing function) are not always the same. Furthermore, when cost savings are pursued and emphasized by the purchasing function, in practice it might mean a lower end user satisfaction (or even dissatisfaction). This further emphasizes the dependence on the service in question. When the purchased service influences the well-being and comfort of the company employees (like cleaning service), it is probably not a good idea to pursue a perceived outcome quality level much lower than 3.00. However, pursuing a perceived outcome quality score close to 5.00 may also not be practical. About the desired level of cleaning service, the external Director of real estate services stated:

"This week I have been visiting in facilities where it (the cleaning service quality) was too good for factory conditions. If you can see your own reflection from a plastic carpet, you don't need that level of quality. And it surely wouldn't result in negative feedback. From a cost-efficient point of view, we would need to find a state where, if there were four smiley faces, that the second best was chosen. Not the best, because that's usually also expensive. It doesn't always mean that though, but normally it (the best quality) means that some extra actions are performed."

When considering a service that is strongly linked to the core competence, for example the maintenance of the production equipment in the case company's case, the acceptable

level of perceived outcome quality should be 4.00 or even higher. In that case, a level as high as possible would be pursued. Furthermore, the operators of those equipment are likely the best evaluators of the quality of the maintenance service, so using the end user satisfaction as a measure of service quality is well justified. The measurement of satisfaction of the end users of the service also supports the purchasing function in serving its own customers better, because in addition to the information about service quality, the purchasing function receives also the information about their customers' satisfaction on the service. Moreover, with minor changes the responsiveness, expertise and perceived outcome quality could be used to measure the service quality of the purchasing function.

The measurement of supplier capability and supplier-customer relationship dimensions raised several implications for the future use of these measurements. The low number of potential respondents in both surveys makes it problematic to compare and link the results with the case company survey results. This same problem is probably present in most of the purchased services: especially, when the purchased service is not linked to the core competence of the buying company, since it probably means that there are not that many supplier employees delivering the service. The low number of responses also limits the statistical analysis that can be used to examine the data. In addition, the differing organizational structures of the supplier and buyer may inhibit the use of the relationship survey in some services. Ideally the respondents to the relationship survey from both companies would be persons who are in contact with each other. Therefore, the appropriate level of analysis should be determined. For the supplier capability and relationship dimensions, the results could be examined at the company level, rather than at unit level. This way, the supplier capability and supplier-customer relationship would be general supplier-specific measures, still enabling the comparison between suppliers of different services.

The suitability of the service quality measures should be assessed individually for each service. The perceived outcome quality can be measured in practically all purchased services, but it is especially useful when the outcome of the service (directly) affects the personnel of the buyer, since it also conveys information about the satisfaction of the personnel. The measurement of process quality (responsiveness and expertise) requires that there is substantial contact between the supplier and the buyer personnel. Therefore, they are best suited for high-involvement and high contact (Chase 1978) services. The use of the process quality dimension in the assessment of cleaning service is justified: only two respondents from the case company noted that the visibility and presence of the cleaning personnel is scarce.

The simplicity and usability of the measurement in practice should also be addressed. Dabholkar et al.'s (2000) results suggest that factors of service quality should be viewed as antecedents rather than components. This implies that when assessing service quality, consumers evaluate different factors affecting the service quality, but they also form a separate overall evaluation of the service quality. Therefore, service quality could also be

assessed by measuring only the overall evaluation of service quality. (Dabholkar et al. 2000, p. 166.) For simplicity, service quality could be measured using only one measure: the overall measure would capture the effects of all the factors related to service quality. The service quality measurement presented in this thesis could be used to develop such a measure.

Measuring supplier capability and the supplier-customer relationship is not meaningful in all purchased services. The measurement of these dimensions of supplier service quality should focus mainly on key suppliers. For example, measuring the quality of the relationship with a supplier that provides office supplies for the company, is probably not feasible. In addition, the size of the supplier has implications for the measurement: when the supplier is very small, it might not be worthwhile to measure the practices of the supplier, if the supplier lacks the resources to properly develop these practices. When measuring supplier capability, the cooperation of the supplier is of utmost importance. It is possible that the supplier does not allow the buyer company to measure its practices in the way that the buyer would want, due to the sensitive nature of the measured issues. This can be influenced to some extent with the factors and survey items, as they can be modified to be more general. It is also important to note that when supplier practices are measured by the buyer company, it is possible that the supplier employees do not answer the questions completely truthfully. This also can be influenced with close collaboration between the buyer and the supplier.

In the future, the viability of using three separate surveys should be assessed. Especially, measuring the supplier-customer relationship may not be viable in the context of service quality, due to the complex nature of the concept. However, measuring the quality of the relationship with key suppliers is still strongly recommended. If the supplier service quality measurement had been conducted by using one common survey for all the respondent groups, the information gathered would have been more limited. On the other hand, using only one survey would have enabled a larger sample size. Basically, responsiveness, expertise and perceived outcome quality factors could have still been used: the case company employees would have assessed how they perceive these aspects, while the supplier employees would have assessed the success of or the ability to perform these aspects from their own point of view. This would have enabled a direct comparison of the results, even at the level of individual survey items. Using one common survey would have prevented gathering information on supplier practices, and therefore also the objective of examining causal relationships. However, in the future the possibility of using only one survey for the measurement should be considered.

Combining the subjective service quality measurement developed in this thesis with objective service quality measures enables a more holistic understanding of service quality. Furthermore, the relative importance of the used measures depends on the service, and should be decided individually for each service. According to Grönroos (2007, p. 76), technical quality of the outcome is normally seen as a prerequisite for good quality: the

technical quality has to be at an acceptable level. However, a good technical quality alone doesn't mean that customers are satisfied with the overall service. For customers to consider total service quality good, also functional (i.e. process) quality has to be good. In some cases, process quality may be more important than outcome quality (Grönroos 1984, p. 42); especially when firms are competing with basically same outcomes, since then it is the functional quality that counts (Grönroos 2007, p. 76). Surprenant & Solomon's (1987, pp. 92-93) results suggest that customer satisfaction may even be more dependent on the process than outcome quality.

When the service quality measurement is applied to other services, most of the items developed here can be used as such. Especially the items in process quality, supplier capability and supplier-customer relationship dimensions are applicable to any service, assuming of course that they are relevant in that service. However, the measurement items in perceived outcome quality should always be chosen and modified according to the service in question to ensure the validity and relevance of the measurement. An important implication of the measurement conducted in this thesis concerns the use of open questions and giving the respondent a possibility to further explain his or her answers. The open comments revealed important and relevant information considering the service and its delivery, that can be used in the development of the service. In many instances the answers to the open questions provided an explanation for the poor score given by the respondent, offering an instant suggestion for development.

One important aspect of the survey analysis was the statistical analysis conducted with the case company data. When the measurement of service quality is applied to other services, the factor analyses should be done again with the data from all the measured services. By doing this, the factor structure can be further confirmed and developed to fit all the measured services. The three-factor structure presented in this thesis was confirmed only in the context of cleaning service, and is not necessarily applicable as such in all services.

Even though the hypothesized model presented in Figure 23. could not be examined in full, the model is however a good starting point for the future development of the supplier service quality measurement. The exploratory and confirmatory factor analyses conducted with the case company data validate the survey items and their use in measuring subjective service quality. This promotes the future use of the survey in the case company. The model for supplier service quality measurement should be examined in full in order to verify the proposed links. In addition, applying the model to other services further strengthens the model and increases the sample size of the measurement data. If and when the links described in the model are verified, development actions can be targeted better, and the service can be developed to better fit the needs of the buying company. This also enables the supplier to better serve the customer. The process and outcome quality dimensions alone convey information about the customer's satisfaction on the service, which is useful also for the supplier.

To reliably examine possible causal relationships between, for example, supplier capability and the perceived service quality, a longitudinal study with multiple data points is required. Also, with data available from a longer period, the trends in the scores of the measured factors can be detected, as well as the effects of development actions on the service quality. As the contact person of the case company noted in a meeting, increasing the data collection frequency could speed up the gathering of a large enough data set: only a part of the survey items would be asked from the respondent at certain times, therefore gradually increasing the amount of data. The key actions points suggested in this section for the future use and development of the supplier service quality measurement are:

- Determining targets for each measure.
- Determining the most appropriate level of analysis for the supplier capability and relationship dimensions, considering the viability and usefulness of the measurement.
- Assessing the viability of using three separate surveys for service quality measurement.
- Assessing the suitability of the developed service quality measures individually for each service.
- Modification of the measurement items when applied to other services, if necessary.
- Confirming the suggested factor structure in other services.
- Further examining and validating the suggested model for supplier service quality measurement.

Based on the research conducted in this thesis, the process for definition and measurement of service quality can be summarized and described. This process along with key issues for each step is presented in Figure 28. The process can also be used as a guide when the service quality measurement is applied to other services.

Figure 28. *The process for service quality definition and measurement with key issues for each step.*

The process presented in Figure 28. has two important implications. First, in defining service quality both the buyer's and the supplier's point of view should be considered. This enables a more comprehensive view on the service and its quality. Furthermore, in constructing a common measurement system, the measures should be decided jointly to ensure the effectiveness of the measurement. Second, the purchasing function should pursue a good comparability of the subjective service quality results. If the measures and measurement items are different even for some of the services, the usability of the measurement suffers due to the decreased amount of comparable data. Furthermore, the maintenance of the measurement system becomes a complicated and time consuming task. Therefore, the subjective measurement of service quality should be constructed in a way that enables the application of the exact same measurement in as many services as possible. The measurement items used in this thesis offer a great starting point.

Finally, the key points of this thesis are summarized in Table 19. The key points are divided into four categories: service quality framework, survey and survey items, service quality measurement, and analysis and interpretation of the measurement results.

Table 19. *The key points of this thesis.*

Category	**Key points**
Service quality framework	• Supplier capability, relationship, and process and outcome quality describe different aspects related to service quality • The four dimensions form a comprehensive view on service quality • Service quality is at least partly dependent on the service and context • Purchasing function's information needs are considered in the framework
Surveys and survey items	• Most of the items used in this research are based on previous studies • The items in the case company survey were validated using factor analyses • The items are generic in nature: most of them can be applied as such to other services • Answers to open questions reveal useful insights about the service
Service quality measurement	• For comparability and simplicity, the measurement should focus on common factors of service quality • Responsiveness, expertise and perceived outcome quality are applicable to almost any service • Measurement of supplier capability and relationship should be applied only with key suppliers • Measuring the factors of supplier capability may be considered as sensitive by the supplier
Analysis and interpretation of the measurement results	• Statistical analysis requires a large enough sample size (> 100) for reliable results • To reliably examine the relationships between constructs requires a longitudinal study • When measuring end-user satisfaction with a five-point Likert scale, a mean score of 3.00 could be considered the lowest acceptable level in any service • Combining subjective and objective measures can provide a more comprehensive view on service quality

The key points provide a brief overview of the main findings of this thesis. The key points aim to summarize the most relevant findings and suggestions of each category.

5.5 Implications of the results on literature

The supplier service quality framework developed in this thesis was developed as a comprehensive description of the aspects of service quality including the viewpoints of both the buyer and the supplier. In service profit chain research the measurement is generally conducted by the service provider. Also in business-to-business service literature the focus has largely been on the supplier's point of view (Jackson et al. 1995, p. 100). The supplier capability dimension of the framework was based on an implication of the service profit chain: the operations and processes of the service supplier can influence service quality. The framework therefore emphasizes the role of both parties in the delivery of quality service, including a larger part of the service delivery process in the measurement. The measurement of the relationship dimension can be seen as a step towards "supplier satisfaction surveys" proposed by Van der Valk & Rozemeijer (2009, p. 7). The idea of supplier satisfaction surveys is to ask the supplier how they perceive the relationship. This way suggestions for improvement are asked from the suppliers. (Van Weele & Rozemeijer 1998, p. 340, 353.)

The results are in line with the suggestion that some service quality dimensions are generic and some are dependent on the service and context (Ladhari 2008, p. 79). This research identified four dimensions of service quality (supplier capability, relationship, and process and outcome quality), which can be considered more or less common for all industrial services. The existing models and conceptualizations of service quality, along with the popular Servqual instrument, were not deemed suitable for the context of this research. This research adopted the perception-based measurement approach to service quality (Cronin & Taylor 1992). The main reasons for this were the criticism that the disconfirmation measurement has received, and the better suitability and simpler implementation of the perception measurement approach.

In this thesis, the surveys and survey items were developed specifically for the purposes of this research. This is an important distinction compared to the service profit chain and service quality literature, where existing survey data is often used. Therefore, the validity of the measurement and the measurement items can be considered to be better. Furthermore, the administration of the surveys could be controlled and managed well compared to the previous studies. When the purpose is to measure service quality for the management and development of the service, it is important that the control over the measurement process is as good as possible, so that the reliability of the results can be ensured.

The results of this research support the finding that service quality is a multidimensional construct (Brady & Cronin 2001, p. 34). Service quality as perceived by the customer can be modelled using Grönroos' (1984) process and outcome quality, which are strongly supported also in the literature (e.g. Brady & Cronin 2001; Gounaris 2005; Ko & Pastore 2005). Furthermore, responsiveness and expertise have been identified also in the

literature as factors of service quality (e.g. Parasuraman et al. 1985; Brady & Cronin 2001), even though the definitions and items of the factors are not the same. Based on the literature review, the existing service quality models (see Appendix H) have generally identified three or four dimensions on which the evaluation of service quality is based on, compared to the two (process and outcome quality) used in this thesis. While the larger number of dimensions might make the service quality evaluation more accurate, the viability of the measurement should always be considered. Especially for practitioners, dividing perceived service quality into process and outcome quality may be sufficient.

The developed supplier service quality model (section 5.2.4.) can be seen as an adaptation of service profit chain (Heskett et al. 1994), with the exception that the measurement was conducted by the buyer company. This had implications on the measurement process, as the buying company collected data concerning the supplier and the supplier employees. Especially, the small number of supplier employees delivering the service hampered the statistical analysis of the results. It was hypothesized that the supplier capability would have an effect on both the process and outcome quality, and that supplier-customer relationship would act as a moderator between supplier capability and process quality. Basically, the suggested model is a simplified service profit chain model, where the supplier's side is represented by supplier's practices and the buyer's side is represented by process and outcome quality, which essentially measure the satisfaction of the customer. As the measurement is conducted from the buying company's point of view, the simplified model could well be justified. However, since the relationship between the supplier employees' and the buyer employees' perceptions of service quality is clearly a complex phenomenon, there is a risk of oversimplifying the measurement.

6. CONCLUSIONS

6.1 Summary of the main findings

This research aimed to develop an approach for supplier service quality measurement in industrial services. The aim was that the measurement could be used jointly by the buyer and supplier in the management and development of the service. The developed supplier service quality measurement was piloted using cleaning service.

First, this research attempted to define service quality in industrial services. As the measurement was to be used by both the buyer and supplier, a broad view on service quality was adopted. Based on existing literature and interviews with the buyer and supplier representatives, service quality was deemed to consist of four dimensions: supplier capability, supplier-customer relationship, process quality and outcome quality. This definition combines the viewpoints of service profit chain (Heskett el al. 1994), relationship quality (e.g. Huntley 2006) and process and outcome quality of the service (Grönroos 1982). Each of the four dimensions in turn consists of a different number of factors: the factors were chosen based on the information needs of the purchasing firm and the supplier. While the dimensions are applicable to any given service, at least some of the factors can be very service-specific. The results are then in line with Ladhari (2008) suggesting that some aspects of quality are common in all services, while some aspects are dependent on the service and context. Furthermore, the relative importance of the dimensions and factors varies depending on the service in question.

Currently, the measurement of cleaning service quality in the case company is dominated by objective measures, such as quality rounds and cleaning frequency. The service quality measurement conducted in this research focused on the subjective side of service quality, because the objective measurement was deemed to be already at a good level. The need to measure the "soft" aspects of purhased services was indicated clearly by the case company. Using both objective and subjective measures, a more comprehensive view on service quality can be achieved. Cleaning service quality was measured using three surveys: the supplier, relationship, and case company surveys. Majority of the survey items were based on existing studies, and a few items were specifically developed for the purposes of this research. In the surveys, a five-point Likert scale was used.

The supplier capability survey measured the current state of the supplier's practices: job enablers, goal clarity, employee empowerment, employee engagement, feedback, and supplier's ability to develop the service. The results suggest that overall the supplier practices are on a very good level in the studied cleaning service. The feedback and supplier's ability to develop the service had the lowest scores, 3.65 and 3.60 respectively. The supplier-customer relationship was measured using two factors: communication and

trust. The results suggest that the supplier respondents assess both the level of communication and trust to be better than do the case company employees. However, the amount of missing data in the relationship questionnaire from the case company's side suggests that the respondents were poorly suited to answer the questions.

For the case company survey data, both exploratory and confirmatory factor analysis was conducted. The analyses resulted in a three-factor model of process and outcome quality: responsiveness, expertise and perceived outcome quality. The exploratory and confirmatory factor analyses also validated the use of the questionnaire items in measuring subjective service quality. Furthermore, most of the items can be used as such also in other services. Overall, the responsiveness of the supplier employees was perceived as good in all production units. The expertise was perceived as very good, apart from Unit 5, where the score was 3.50. Therefore, the results suggest that the case company employees are generally satisfied with the supplier employees delivering the service. However, the results on perceived outcome quality suggest that the case company employees were not satisfied with the cleaning service of the supplier, the mean of all the Units being only 3.17. The comments to the open questions further elaborated that the schedule of cleaning was perceived to be too strict, as the supplier employees do not have enough time to perform the cleaning service satisfactorily.

An important aspect in the suggested measurement system is the viability and simplicity of the measurement. Three separate surveys with different respondent groups may cause the measurement process to be too complex. Especially, the inclusion of supplier-customer relationship quality in the measurement of service quality should be assessed, due to the complexity of the construct. The quality of the relationship should however be measured in some way, but as with the supplier capability, the measurement should focus only on key suppliers in order to ensure the effectiveness and profitability of the measurement.

This thesis also developed a model of supplier service quality, linking supplier capability, supplier-customer relationship and process and outcome quality. Due to the small sample sizes of the supplier and relationship questionnaires, only the process and outcome quality were examined. The relationships between responsiveness, expertise and perceived outcome quality were explored using correlation analysis, structural equation modeling and path modeling. The results indicated that there is a connection between responsiveness, expertise and perceived outcome quality, i.e. the process and outcome quality of the service. Responsiveness and expertise were closely related, which was expected, since they were hypothesized to measure the same higher order construct, process quality. Based on the results, expertise had a greater effect on perceived outcome quality than responsiveness. Both of these relationships were statistically significant in the path modeling. Overall, the results provided support for the use of process and outcome quality in measuring subjective service quality.

6.2 Managerial implications

This study offers several useful insights for managers concerning supplier service quality and its measurement. Even though this research was conducted as a case study, the undergone process reported in this thesis can be generally useful to managers in various organizations. An important implication of this research is the context specific nature of service quality: a universal definition of service quality does not yet exist. It is quite clear that the definition of service quality differs at least in some aspects depending on the service in question. Even if the dimensions and factors of service quality are deemed to be identical in two different services, the relative importance of these dimensions and factors may differ. Therefore, managers should always consider the particular features of a service when aiming to define and measure the service quality. However, managers should keep in mind that using service-specific measures inhibits the comparison of service quality between services. In addition, more uniform data is obtained for statistical analysis by using the same measures in all services, enabling better reliability of the measurement.

The results suggest that the purchased service quality measurement is dominated by "hard" objective measures, one of which are the costs, since these are relatively easy to measure. The "soft" subjective side of the measurement is lacking, even though it is needed in forming a comprehensive understanding of the service quality. In many services, objective measures alone fail to convey relevant information about the actual success of the service. Using end user (customer) satisfaction as measure of service quality can provide valuable information about the service.

This research aimed to develop a general framework for industrial service quality. The developed framework offers a great starting point for managers aiming to determine the important factors of a service. Especially, the service quality framework highlights both the buyer's and the supplier's part in service quality. It is important that the supplier and the supplier employees are involved in the discussion on service quality. First, this provides for an expert opinion on the service and its delivery, and second, it opens a communication channel that is needed for the efficient management and development of the service.

The measurement of supplier service quality and the analysis of the results offer important implications for managers. This thesis provides support for the use of a survey in service quality measurement. However, the length and complexity of the survey is critical for the respondent's willingness to respond, and for the ease of administering and maintenance of the survey. Even though not empirically tested, using the perception-only measurement relative to the disconfirmation is arguably more suitable for the measurement of service quality conducted by practitioners, as the perception approach is easier and simpler to implement. The surveys in this thesis were developed for the specific purpose of supplier service quality measurement, and can be used as a starting point in developing a service

quality measurement in any organization. Especially, the case company survey and the respective items were validated through statistical analysis, and the survey is applicable to other services, therefore enabling the comparison of supplier service quality between services.

6.3 Limitations and criticism

The research conducted in this thesis is not without limitations. Saunders et al. (2009, pp. 156-158) suggest that the credibility of the research findings needs to be assessed in terms of reliability, validity and generalizability. The research in this thesis consisted of two phases: the determination and the measurement of supplier service quality. The determination phase was qualitative in nature, as semi-structured interviews and literature review were the main source of data. The actual measurement of cleaning service quality was predominantly quantitative: the results are largely based on statistical analysis. Therefore, the limitations are different for these two phases.

This thesis developed a general framework for supplier service quality using existing literature and semi-structured interviews. The reliability of the interview results can be considered good. In the interviews the structure was the same and the same main question were asked from all the interviewees. Furthermore, the interview questions were carefully designed in accordance with the research questions. If conducted again, the results would probably be more or less the same. However, some additional aspects and factors of service quality might be discovered, due to the subjective nature of service quality. The validity of the interview results is good. The sampling for the interviews was purposive: the interviewees were chosen in collaboration with the case company and supplier contact persons. Therefore, it can be argued that the right people were obtained for the interviews. The interview results are not generalizable as such to all companies, but can be considered well representative in the context of industrial companies: the aspects of purchasing and the factors of quality discussed in the interviews are most likely relevant in other companies also. The developed framework for supplier service quality serves its intended purpose well: a general framework to assist in determining the relevant aspects of supplier quality in industrial services.

Based on the framework, the measurement of service quality was designed using three surveys for different respondent groups. The final versions of the surveys and chosen quality factors as well as the respective items were approved by the case company representatives. The surveys were specifically designed and developed for the purposes of this research, and most of the items were based on existing studies. The item selection was naturally directed to studies that reported the used survey items. Therefore, it can be argued that the validity of the supplier service quality measurement is good. Furthermore, the case company survey items were further validated by the statistical analysis of survey data. A large part of the results of the supplier service quality surveys are a result of

different statistical analyses. The reliability and significance of the results of the applied methods have been reported and addressed in this thesis with the results.

Considering the small sample size of the surveys, the results of the statistical analyses and survey results should be interpreted with caution. Especially, the small sample size arguably has an effect on the reliability of the statistical analyses conducted in this thesis. Considering the supplier capability and supplier-customer relationship surveys, the sample sizes were too small to make reliable conclusions from the results, and the results could not be examined at the intended Unit level. In the relationship survey the non-correspondence of the respondents also limited the analysis. In retrospect, the requirements of the statistical analyses used in this thesis should have been taken better into account already in the beginning of the research. For example, the sample size of the surveys should have been at least 100, and preferably over 200, in order to increase the validity of the results. However, in the case of the supplier employees, it is most likely not possible to obtain a large sample size. This is due to fact that when the viewpoint of the customer is adopted, the focus immediately shifts to those supplier employees that deliver the service to this particular customer. Therefore, there probably is not that many supplier employees delivering the service in the first place, regardless of the purchased service.

In the literature, the sample sizes in service profit chain studies are often over a thousand responses, covering data collection from several years. The time constraints of the research affected the execution and the data collection of this research. To reliably examine the links suggested in the supplier service quality measurement model, a longitudinal study is required. Furthermore, when using survey data in the analysis, common method bias must be addressed. Ideally, multiple sources in the data collection are used (Chang et al. 2010, p. 182).

The results of the supplier service quality measurement were obtained from a specific group of people from specific companies considering a specific service at a specific time. However, the reliability and the validity of the case company survey results can be considered good. Also, the repeatability of the survey and statistical analysis of this research is good. If conducted again, the results on cleaning service quality would probably be quite similar, especially if the time between the measurements was not long. The generalizability of the case company survey can also be considered good: it can be used to measure the subjective quality of cleaning service in almost any industrial organization. Furthermore, it can be used to measure the service quality of practically any industrial service. It is notable, that the factor structure was only confirmed in the case of cleaning service, and that no alternative structures, other than the initial four-factor structure, were tested. Therefore, the relevance of the responsiveness, expertise and perceived outcome quality should always be determined for each service separately.

6.4 Implications for future research

This research adds to the vast literature of service quality by offering an approach for supplier service quality measurement. The framework developed in this thesis is meant to be widely applicable to different industrial services. In the literature, there is no consensus on the definition of service quality that is generally applicable to all services. However, finding common quality dimensions and factors in different services can be useful, because it enables the comparison of services in those aspects. To find common aspects and measures in services, service classifications, such as the one provided by Chase (1978), could be used. As in process and outcome quality, also in supplier capability and supplier-customer relationship the focus should be in finding the generic factors that have the most effect on service quality. In the future, examining the relative importance of the service quality dimensions and factors in different services may also be useful, as it reveals the most critical aspects of a service.

In the service profit chain and service quality literature, the measurement is almost invariably conducted from the viewpoint of the company that delivers the service. Therefore, the customer's actual needs may not be that well present in the measurement. In this research, the service quality is examined from the purchaser's point of view. In the case of service purchasing this is an obvious choice, because the purchasing function needs to ensure that the purchased service creates value for the buying company. However, the lack of research in service quality measurement from the customer's point of view is striking. By taking the customer's perspective, it might be possible to better address the customer's needs. Ultimately, considering the management and development of the service, both the buyer and supplier organizations should be considered when it comes to service quality.

The suggested model for supplier service quality measurement could not be examined in full, but the model offers an interesting concept for service quality management. Examining the relationships between the supplier's practices and service quality could enable a more effective management and development of a purchased service. The developed model should be verified in different services using a longitudinal study to see whether the measured constructs and links prove to be lasting over time. As per service profit chain literature, the supplier service quality model could also ultimately be extended to include organizational outcomes, such as profit and productivity, of the buying company.

REFERENCES

Anderson, E.W. & Mittal, V. (2000). Strengthening the satisfaction-profit chain, Journal of Service Research, Vol. 3(2), pp. 107-120.

Anderson, J.C. & Narus, J.A. (1990). A model of distributor firm and manufacturer firm working partnerships, Journal of Marketing, pp. 42-58.

Armstrong, J.S. & Overton, T.S. (1977). Estimating nonresponse bias in mail surveys, Journal of Marketing Research, Vol. 14(3), pp. 396-402.

Asubonteng, P., McCleary, K.J. & Swan, J.E. (1996). SERVQUAL revisited: a critical review of service quality, Journal of Services marketing, Vol. 10(6), pp. 62-81.

Axelsson, B. & Wynstra, F. (2002). Buying business services, John Wiley, Chichester, 282 p.

Bettencourt, L.A. & Brown, S.W. (1997). Contact employees: relationships among workplace fairness, job satisfaction and prosocial service behaviors, Journal of Retailing, Vol. 73(1), pp. 39-61.

Bhagwat, R. & Sharma, M.K. (2007). Performance measurement of supply chain management: a balanced scorecard approach, Computers & Industrial Engineering, Vol. 53(1), pp. 43-62.

Bitner, M.J., Booms, B.H. & Mohr, L.A. (1994). Critical service encounters: the employee's viewpoint, Journal of Marketing, Vol. 58(4), pp. 95-106.

Bitner, M.J. (1990). Evaluating service encounters: the effects of physical surroundings and employee responses, Journal of Marketing, Vol. 54(2), pp. 69-82.

Bitner, M.J. (1992). Servicescapes: the impact of physical surroundings on customers and employees, Journal of Marketing, Vol. 56(2), pp. 57-71.

Boshoff, C. & Leong, J. (1998). Empowerment, attribution and apologising as dimensions of service recovery: an experimental study, International Journal of Service Industry Management, Vol. 9(1), pp. 24-47.

Boshoff, C. (1999). RECOVSAT: an instrument to measure satisfaction with transaction-specific service recovery, Journal of Service Research, Vol. 1(3), pp. 236-249.

Bowen, D.E. (2008). Linking Various Perspectives on Service, in: Smith, D.B. (ed.), The people make the place: dynamic linkages between individuals and organizations, CRC Press, pp. 157-174.

Bowen, D.E., Gilliland, S.W. & Folger, R. (2000). HRM and service fairness: how being fair with employees spills over to customers, Organizational dynamics, Vol. 27(3), pp. 7-23.

Brady, M.K. & Cronin Jr, J.J. (2001). Some new thoughts on conceptualizing perceived service quality: a hierarchical approach, Journal of Marketing, Vol. 65(3), pp. 34-49.

Brady, M.K., Cronin, J.J. & Brand, R.R. (2002). Performance-only measurement of service quality: a replication and extension, Journal of Business Research, Vol. 55(1), pp. 17-31.

Brady, M.K., Voorhees, C.M., Cronin Jr, J.J. & Bourdeau, B.L. (2006). The good guys don't always win: the effect of valence on service perceptions and consequences, Journal of Services Marketing, Vol. 20(2), pp. 83-91.

Brown, T.J., Churchill, G.A. & Peter, J.P. (1993). Improving the measurement of service quality, Journal of Retailing, Vol. 69(1), pp. 127-139.

Caro, L.M. & García, J.A.M. (2008). Developing a multidimensional and hierarchical service quality model for the travel agency industry, Tourism Management, Vol. 29(4), pp. 706-720.

Caro, L.M. & Garcia, J.A.M. (2007). Measuring perceived service quality in urgent transport service, Journal of Retailing and Consumer Services, Vol. 14(1), pp. 60-72.

Carter, C.R. & Ellram, L.M. (2003). Thirty-five years of the Journal of Supply Chain Management: where have we been and where are we going? Journal of Supply Chain Management, Vol. 39(1), pp. 27-39.

Chang, S., Van Witteloostuijn, A. & Eden, L. (2010). From the editors: common method variance in international business research, Journal of International Business Studies, Vol. 41(2), pp. 178-184.

Cook, D.P., Goh, C. & Chung, C.H. (1999). Service typologies: a state of the art survey, Production and Operations Management, Vol. 8(3), pp. 318-338.

Cousins, P.D., Lawson, B. & Squire, B. (2008). Performance measurement in strategic buyer-supplier relationships: the mediating role of socialization mechanisms, International Journal of Operations & Production Management, Vol. 28(3), pp. 238-258.

Cronin, J.J., Brady, M.K. & Hult, G.T.M. (2000). Assessing the effects of quality, value, and customer satisfaction on consumer behavioral intentions in service environments, Journal of Retailing, Vol. 76(2), pp. 193-218.

Dabholkar, P.A., Thorpe, D.I. & Rentz, J.O. (1996). A measure of service quality for retail stores: scale development and validation, Journal of the Academy of Marketing Science, Vol. 24(1), pp. 3-16.

Day, M. & Lichtenstein, S. (2006). Strategic supply management: the relationship between supply management practices, strategic orientation and their impact on organisational performance, Journal of Purchasing and Supply Management, Vol. 12(6), pp. 313-321.

De Winter, J., Dodou, D. & Wieringa, P. (2009). Exploratory factor analysis with small sample sizes, Multivariate Behavioral Research, Vol. 44(2), pp. 147-181.

DeWitt, T., Nguyen, D.T. & Marshall, R. (2008). Exploring customer loyalty following service recovery: the mediating effects of trust and emotions, Journal of Service Research, Vol. 10(3), pp. 269-281.

Doney, P.M. & Cannon, J.P. (1997). An examination of the nature of trust in buyer-seller relationships, the Journal of Marketing, Vol. 61(2), pp. 35-51.

Durvasula, S., Lysonski, S. & Mehta, S.C. (1999). Testing the SERVQUAL scale in the business-to-business sector: the case of ocean freight shipping service, Journal of Services Marketing, Vol. 13(2), pp. 132-150.

Dyer, J.H. & Singh, H. (1998). The relational view: cooperative strategy and sources of interorganizational competitive advantage, Academy of Management Review, Vol. 23(4), pp. 660-679.

Ellram, L.M. (1991). Key success factors and barriers in international purchasing partnerships, Management Decision, Vol. 29(7), pp. 38-44.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C. & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods, Vol. 4(3), pp. 272-299.

Fitzsimmons, J.A., Noh, J. & Thies, E. (1998). Purchasing business services, Journal of Business & Industrial Marketing, Vol. 13(4/5), pp. 370-380.

Garvin, D.A. (1988). Managing quality: the strategic and competitive edge, Free Press, New York, 319 p.

Gelade, G.A. & Ivery, M. (2003). The impact of human resource management and work climate on organizational performance, Personnel Psychology, Vol. 56(2), pp. 383-404.

Gelade, G.A. & Young, S. (2005). Test of a service profit chain model in the retail banking sector, Journal of Occupational and Organizational Psychology, Vol. 78(1), pp. 1-22.

Gerbing, D.W. & Hamilton, J.G. (1996). Viability of exploratory factor analysis as a precursor to confirmatory factor analysis, Structural Equation Modeling: A Multidisciplinary Journal, Vol. 3(1), pp. 62-72.

Giannakis, M. (2007). Performance measurement of supplier relationships, Supply Chain Management: An International Journal, Vol. 12(6), pp. 400-411.

Gounaris, S. (2005). Measuring service quality in b2b services: an evaluation of the SERVQUAL scale vis-à-vis the INDSERV scale, Journal of Services Marketing, Vol. 19(6), pp. 421-435.

Grönroos, C. (1984). A service quality model and its marketing implications, European Journal of Marketing, Vol. 18(4), pp. 36-44.

Grönroos, C. (1988). Service quality: the six criteria of good perceived service, Review of Business, Vol. 9(3), pp. 10-13.

Grönroos, C. (2007). Service management and marketing: customer management in service competition, 3rd ed., John Wiley & Sons Ltd, Chichester, 483 p.

Grönroos, C. (1982). Strategic management and marketing in the service sector, Swedish School of Economics and Business Administration, Helsingfors, 222 p.

Grover, V., Cheon, M.J. & Teng, J.T. (1996). The effect of service quality and partnership on the outsourcing of information systems functions, Journal of Management Information Systems, Vol. 12(4), pp. 89-116.

Gunasekaran, A., Patel, C. & Tirtiroglu, E. (2001). Performance measures and metrics in a supply chain environment, International Journal of Operations & Production Management, Vol. 21(1/2), pp. 71-87.

Gunasekaran, A., Patel, C. & McGaughey, R.E. (2004). A framework for supply chain performance measurement, International Journal of Production Economics, Vol. 87(3), pp. 333-347.

Hahn, C.K., Watts, C.A. & Kim, K.Y. (1990). The supplier development program: a conceptual model, Journal of Supply Chain Management, Vol. 26(2), pp. 2-7.

Hair, J.F., Anderson, R.E., Babin, B.J. & Black, W.C. (2010). Multivariate data analysis: a global perspective, Pearson Upper Saddle River, NJ, 784 p.

Hallowell, R., Schlesinger, L.A. & Zornitsky, J. (1996). Internal service quality, customer and job satisfaction: linkages and implications for management, People and Strategy, Vol. 19(2), pp. 20-31.

Harmon, J., Scotti, D.J., Behson, S. & Farias, G. (2003). Effects of high-involvement work systems on employee satisfaction and service costs in veterans healthcare, Journal of Healthcare Management, Vol. 48(6), pp. 393-406.

Hart, C.W.L., Heskett, J.L. & Sasser Jr., W.E. (1990). The profitable art of service recovery, Harvard Business Review, Vol. 68(4), pp. 148-156.

Haywood-Farmer, J. (1988). A conceptual model of service quality, International Journal of Operations & Production Management, Vol. 8(6), pp. 19-29.

Heikkilä, J., Vuori, M., & Laine, J.A.T., (2013). Purchasing business services, Teknologiainfo Teknova, Helsinki, 182 p.

Hellier, P.K., Geursen, G.M., Carr, R.A. & Rickard, J.A. (2003). Customer repurchase intention: a general structural equation model, European Journal of Marketing, Vol. 37(11/12), pp. 1762-1800.

Heskett, J.L., Jones, T.O., Loveman, G.W., Sasser, W.E. & Schlesinger, L.A. (1994). Putting the service-profit chain to work, Harvard Business Review, Vol. 72(2), pp. 164-174.

Holmlund, M. & Kock, S. (1995). Buyer perceived service quality in industrial networks, Industrial Marketing Management, Vol. 24(2), pp. 109-121.

Homburg, C. & Garbe, B. (1999). Towards an improved understanding of industrial services: quality dimensions and their impact on buyer-seller relationships, Journal of Business-to-Business Marketing, Vol. 6(2), pp. 39-71.

Homburg, C., Wieseke, J. & Hoyer, W.D. (2009). Social identity and the service-profit chain, Journal of Marketing, Vol. 73(2), pp. 38-54.

Huntley, J.K. (2006). Conceptualization and measurement of relationship quality: linking relationship quality to actual sales and recommendation intention, Industrial Marketing Management, Vol. 35(6), pp. 703-714.

Iloranta, K., & Pajunen-Muhonen, H., (2015). Hankintojen johtaminen: ostamisesta toimittajamarkkinoiden hallintaan, 4th ed., Tietosanoma, Helsinki, 427 p.

Jackson, R.W., Neidell, L.A. & Lunsford, D.A. (1995). An empirical investigation of the differences in goods and services as perceived by organizational buyers, Industrial Marketing Management, Vol. 24(2), pp. 99-108.

Jiang, Z., Shiu, E., Henneberg, S. & Naude, P. (2016). Relationship quality in business to business relationships—reviewing the current literatures and proposing a new measurement model, Psychology & Marketing, Vol. 33(4), pp. 297-313.

Kaplan, R.S. & Norton, D.P. (1992). The balanced scorecard - measures that drive performance, Harvard Business Review, Vol. 70(1), pp. 71-79.

Ko, Y.J. & Pastore, D.L. (2005). A hierarchical model of service quality for the recreational sport industry, Sport Marketing Quarterly, Vol. 14(2), pp. 84-97.

Korpela, T. (2015). Dyadic relations between the main contractor and its suppliers: a case study to clarify critical factors in Metso Paper Ltd, Doctoral Dissertation, Jyväskylä studies in humanities 247, 271 p.

Krause, D.R. & Scannell, T.V. (2002). Supplier development practices: product-and service-based industry comparisons, Journal of Supply Chain Management, Vol. 38(1), pp. 13-21.

Kumar, N., Scheer, L.K. & Steenkamp, J.E. (1995a). The effects of perceived interdependence on dealer attitudes, Journal of Marketing Research, Vol. 32(3), pp. 348-356.

Kumar, N., Scheer, L.K. & Steenkamp, J.E. (1995b). The effects of supplier fairness on vulnerable resellers, Journal of Marketing Research, Vol. 32(1), pp. 54-65.

Levitt, T. (1972). Production-line approach to service, Harvard Business Review, Vol. 50(5), pp. 41-52.

Lovelock, C.H. (1983). Classifying services to gain strategic marketing insights, Journal of Marketing, Vol. 47(3), pp. 9-20.

Lovelock, C. & Gummesson, E. (2004). Whither services marketing? In search of a new paradigm and fresh perspectives, Journal of Service Research, Vol. 7(1), pp. 20-41.

Marquardt, K., Olaru, M. & Ceausu, I. (2017). Study on the development of quality measurements models for steering business services in relation to customer satisfaction, Amfiteatru Economic, Vol. 19(44), pp. 95-109.

Metsä Fibre Annual review 2015, web page. Available (accessed 25.8.2106): http://www.metsafibre.com/en/Documents/Publications/Metsa-Fibre-Annual-Review-2015.pdf.

Metsä Group About us, web page. Available (accessed 25.08.2016): http://www.metsagroup.com/en/about-us/Pages/default.aspx.

Metsä Group Purchasing at Metsä Group, web page. Available (accessed 25.8.2016): http://www.metsagroup.com/en/Contact/purchasing/Pages/default.aspx.

Michel, S., Bowen, D. & Johnston, R. (2009). Why service recovery fails: tensions among customer, employee, and process perspectives, Journal of Service Management, Vol. 20(3), pp. 253-273.

Min, H., Chia, A., Goh, M. & Hum, S. (2009). Performance measurement in supply chain entities: balanced scorecard perspective, Benchmarking: An International Journal, Vol. 16(5), pp. 605-620.

Neely, A., Gregory, M. & Platts, K. (1995). Performance measurement system design: a literature review and research agenda, International Journal of Operations & Production Management, Vol. 15(4), pp. 80-116.

Osborne, J.W. & Costello, A.B. (2009). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis, Pan-Pacific Management Review, Vol. 12(2), pp. 131-146.

Parasuraman, A., Zeithaml, V.A. & Berry, L.L. (1985). A conceptual model of service quality and its implications for future research, Journal of Marketing, Vol. 49(4), pp. 41-50.

Parasuraman, A., Berry, L.L. & Zeithaml, V.A. (1991). Refinement and reassessment of the SERVQUAL scale, Journal of Retailing, Vol. 67(4), pp. 420-450.

Parasuraman, A., Zeithaml, V.A. & Berry, L.L. (1994). Alternative scales for measuring service quality: a comparative assessment based on psychometric and diagnostic criteria, Journal of Retailing, Vol. 70(3), pp. 201-230.

Parasuraman, A., Zeithaml, V.A. & Berry, L.L. (1988). Servqual: a multiple-item scale for measuring consumer perceptions of service quality, Journal of Retailing, Vol. 64(1), pp. 12-40.

Podsakoff, P.M., MacKenzie, S.B., Lee, J. & Podsakoff, N.P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. Journal of Applied Psychology, Vol. 88(5), pp. 879-903.

Pugh, S.D., Dietz, J., Wiley, J.W. & Brooks, S.M. (2002). Driving service effectiveness through employee-customer linkages, The Academy of Management Executive, Vol. 16(4), pp. 73-84.

Ryan, A., Schmit, M. & Johnson, R. (1996). Attitudes and effectiveness: examining relations at an organizational level, Personnel Psychology, Vol. 49(4), pp. 853-882.

Salanova, M., Agut, S. & Peiró, J.M. (2005). Linking organizational resources and work engagement to employee performance and customer loyalty: the mediation of service climate. Journal of Applied Psychology, Vol. 90(6), pp. 1217-1227.

Saunders, M., Lewis, P. & Thornhill, A. (2009). Research methods for business students, 5th ed., Pearson Education, 586 p.

Schmidt, M.J. & Hollensen, S. (2006). Marketing research: an international approach, Pearson education, 632 p.

Schneider, B. & Bowen, D.E. (1993). The service organization: human resources management is crucial, Organizational Dynamics, Vol. 21(4), pp. 39-52.

Schneider, B., Ehrhart, M.G., Mayer, D.M., Saltz, J.L. & Niles-Jolly, K. (2005). Understanding organization-customer links in service settings, Academy of Management Journal, Vol. 48(6), pp. 1017-1032.

Schneider, B., Hanges, P.J., Smith, D.B. & Salvaggio, A.N. (2003). Which comes first: employee attitudes or organizational financial and market performance? Journal of Applied Psychology, Vol. 88(5), pp. 836-851.

Schneider, B., Parkington, J.J. & Buxton, V.M. (1980). Employee and customer perceptions of service in banks, Administrative Science Quarterly, pp. 252-267.

Schneider, B., White, S.S. & Paul, M.C. (1998). Linking service climate and customer perceptions of service quality: tests of a causal model. Journal of Applied Psychology, Vol. 83(2), pp. 150-163.

Schweikhart, S.B., Strasser, S. & Kennedy, M.R. (1993). Service recovery in health services organizations, Journal of Healthcare Management, Vol. 38(1), pp. 3-21.

Sergeant, A. & Frenkel, S. (2000). When do customer contact employees satisfy customers? Journal of Service Research, Vol. 3(1), pp. 18-34.

Seth, N., Deshmukh, S. & Vrat, P. (2006). SSQSC: a tool to measure supplier service quality in supply chain, Production Planning & Control, Vol. 17(5), pp. 448-463.

Seth, N., Deshmukh, S. & Vrat, P. (2005). Service quality models: a review, International Journal of Quality & Reliability Management, Vol. 22(9), pp. 913-949.

Sheth, J.N. (1996). Organizational buying behavior: past performance and future expectations, Journal of Business & Industrial Marketing, Vol. 11(3/4), pp. 7-24.

Smeltzer, L.R. & Ogden, J.A. (2002). Purchasing professionals' perceived differences between purchasing materials and purchasing services, Journal of Supply Chain Management, Vol. 38(1), pp. 54-70.

Spina, G., Caniato, F., Luzzini, D. & Ronchi, S. (2013). Past, present and future trends of purchasing and supply management: an extensive literature review, Industrial Marketing Management, Vol. 42(8), pp. 1202-1212.

Spreitzer, G.M. (1995). Psychological empowerment in the workplace: dimensions, measurement, and validation, Academy of Management Journal, Vol. 38(5), pp. 1442-1465.

Stevenson, W.J. (2009). Operations management, 10th ed., McGraw-Hill/Irwin, Boston (MA), 906 p.

Surprenant, C.F. & Solomon, M.R. (1987). Predictability and personalization in the service encounter, Journal of Marketing, Vol. 51(2), pp. 86-96.

Talluri, S. & Sarkis, J. (2002). A model for performance monitoring of suppliers, International Journal of Production Research, Vol. 40(16), pp. 4257-4269.

Teas, R.K. (1993). Expectations, performance evaluation, and consumers' perceptions of quality, Journal of Marketing, Vol. 57(4), pp. 18-34.

Tornow, W.W. & Wiley, J.W. (1991). Service quality and management practices: a look at employee attitudes, customer satisfaction, and bottom-line consequences, People and Strategy, Vol. 14(2), pp. 105.

Ulaga, W. & Eggert, A. (2006). Value-based differentiation in business relationships: gaining and sustaining key supplier status, Journal of Marketing, Vol. 70(1), pp. 119-136.

Van Weele, A.J. & Rozemeijer, F.A. (1998). Mirror, mirror on the wall... Let suppliers guide you towards improvement, in: Capaldo, G., Esposito, E., Lo Sorto, C. & Raffa, M. (ed.), Supply Management, Edizioni Scientifiche Italiane, Napoli, pp. 337-353.

Van Weele, A.J. (2014). Purchasing and supply chain management: analysis, strategy, planning and practice, 6th ed., Cengage Learning EMEA, 438 p.

Vargo, S.L. & Lusch, R.F. (2004a). Evolving to a new dominant logic for marketing, Journal of Marketing, Vol. 68(1), pp. 1-17.

Vargo, S.L. & Lusch, R.F. (2004b). The four service marketing myths remnants of a goods-based, manufacturing model, Journal of Service Research, Vol. 6(4), pp. 324-335.

Voss, C., Tsikriktsis, N. & Frohlich, M. (2002). Case research in operations management, International Journal of Operations & Production Management, Vol. 22(2), pp. 195-219.

Zeithaml, V.A., Parasuraman, A. & Berry, L.L. (1985). Problems and strategies in services marketing, Journal of Marketing, Vol. 49(2), pp. 33-46.

Zeithaml, V.A., Parasuraman, A. & Berry, L.L. (1990). Delivering quality service: balancing customer perceptions and expectations, Free Press, New York, 226 p.

Zemke, R. & Bell, C. (1990). Service recovery: doing it right the second time, Training: The Magazine of Human Resources Development, Vol. 27(6), pp. 42-48.

Zimmermann, F. & Foerstl, K. (2014). A meta-analysis of the "purchasing and supply management practice–performance link", Journal of Supply Chain Management, Vol. 50(3), pp. 37-54

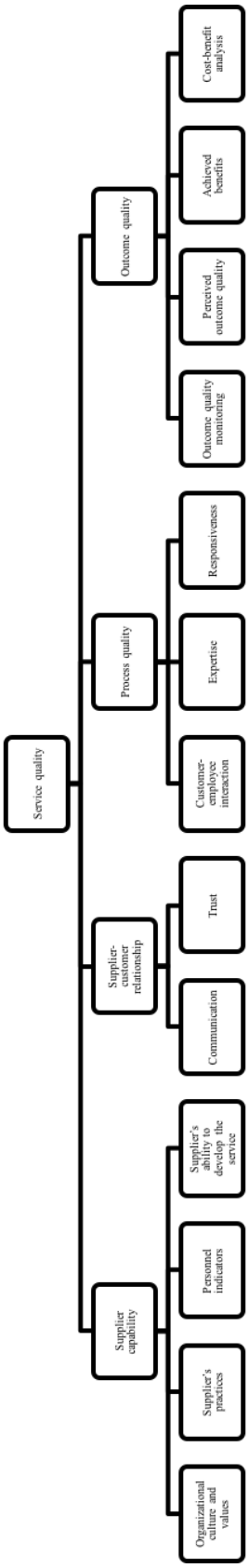
APPENDIX A: INTERVIEW QUESTIONS FOR THE CASE COMPANY REPRESENTATIVES

1. Your name and responsibilities at Metsä Group?
2. What are the information needs of Metsä Group concerning industrial services and their quality?
    a. What kind of information is needed about services?
        i. Especially from the purchasing point of view?
    b. Are there specific information needs concerning cleaning service?
    c. What are the main purposes of this information?
    d. Do you want to say something else about information needs?
3. How would you define the quality of industrial services?
    a. From which dimensions does it consist of?
    b. From your point of view, what are the most important things in quality? Why?
4. What aspects in the supplier's operations and activities or in the relationship affect service quality?
    a. Are there some aspects in the supplier's operations that you see as important in ensuring the delivery of successful service?
    b. Would it be beneficial to get information about these aspects from the supplier?
    c. Is there some other information that you think would be useful to get from the supplier, regarding services and their quality?
5. How is the performance of services monitored?
    a. What are the most important measures in industrial services? What about cleaning services?
    b. How are these results used?
        i. Are they discussed with the supplier?
    c. Is some measure missing, that you think should be measured?
6. Presenting and reviewing of the developed supplier service quality framework (provided at the interview and also beforehand).
    a. What do you think about the framework?
        i. Is the structure of the framework reasonable?
        ii. Do any of the dimensions of factors need further specification?
    b. Does the framework provide relevant information concerning the information needs of purchasing?
        i. Is something relevant missing?
    c. Does the framework adequately describe service quality?
        i. What do you think about the dimensions?
        ii. Should some dimensions(s) be added or removed?
    d. What do you think about the factors?
        i. Should some factor(s) be added or removed?
    e. Would you develop this framework in any way?
        i. How?
7. Are there any other thoughts or comments about information needs, service quality, the supplier service quality framework or measurement in general?

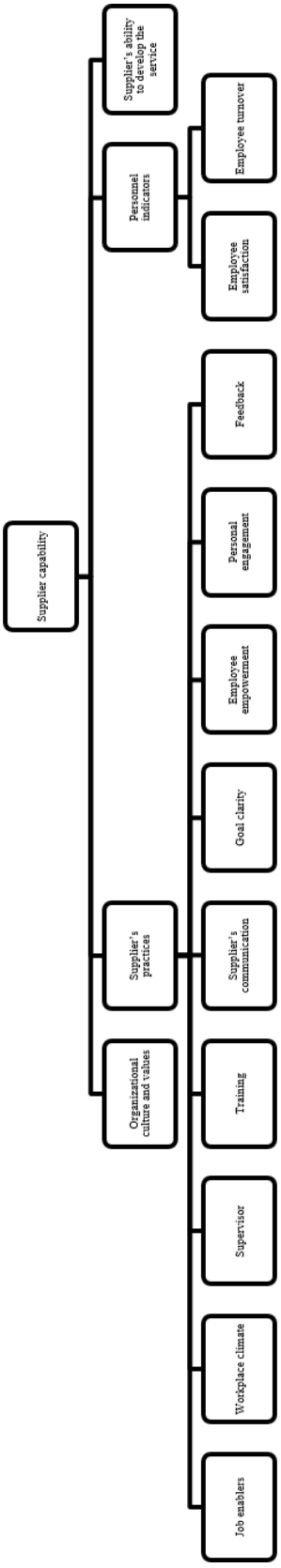
APPENDIX B: INTERVIEW QUESTIONS FOR THE SUPPLIER REPRESENTATIVES

1. Your name and responsibilities at your company?
2. What are the information needs of your company concerning industrial services and their quality at the facilities of Metsä Group?
    a. How does your company define the quality of its services (in relation to, e.g. Grönroos' quality definition)?
        i. Especially in cleaning services but also in e.g. customer relationships?
    b. How does your company measure the quality and performance of its services?
        i. Is there some rationale for using the presented measures?
        ii. Are there some aspects in service quality that you would like to include to the measurement?
        iii. (How do you ensure that the wanted outcome is achieved?)
    c. How often is this information gathered?
    d. How and to what extent is this information used?
        i. Are there some information need not satisfied with the current measurement?
    e. Are the results communicated to Metsä Group?
        i. If not, then would it be possible?
3. What factors are important in the delivery of a successful cleaning service?
    a. What about especially in Metsä Group's case? Are there some special features dependent on the specific customer?
    b. How can you influence these things?
    c. (Vision: "We are going to be the world's greatest service organization", what does this mean in practice?)
4. How are different levels of (cleaning) service defined, i.e. what options does the buyer have when purchasing cleaning service?
    a. How do you ensure that the wanted outcome is achieved?
5. Presenting and reviewing of the developed supplier service quality framework (provided at the interview and also beforehand).
    a. Are the factors in the framework relevant concerning service quality?
        i. Is something relevant missing?
    b. Do you measure the factors related to the Supplier's own processes in the framework?
        i. Is this information shared with customers?
        ii. Do you think it would be beneficial to share this information?
    c. What do you think about the framework?
        i. Is the structure of the model reasonable?
        ii. In your opinion, how well does the framework describe service quality?
6. Are there any other thoughts or comments about information needs, service quality, the supplier service quality framework or measurement in general?

APPENDIX C: SUPPLIER SERVICE QUALITY FRAMEWORK



APPENDIX D: SUPPLIER CAPABILITY DIMENSION AND THE RELATED FACTORS



APPENDIX E: SURVEY ITEMS FOR THE SUPPLIER'S EMPLOYEES (SUPPLIER CAPABILITY)

Coding	Factor and items	Original source
	Job enablers	
S1	I have sufficient acquirements (knowledge and skills) to do my job well	Schneider et al. 2003
S2	Working conditions at Metsä Group's facilities allow me to be about as productive as I could be	Harmon et al. 2003
S3	I'm satisfied with my physical working conditions (e.g. the break rooms, working environment)	Schneider et al. 2003
S4	Overall, I'm satisfied with my ability to meet customers' needs	Sergeant & Frenkel 2000
	Goal clarity	
S5	I understand the key strategic objectives of <supplier name>	Gelade & Young 2005
S6	I am well aware of my working community's objectives	Fischer 2012
S7	I know the key indicators for my working community	Fischer 2012
S8	I know very well the objectives set for me	Fischer 2012
S9	I can perform the tasks assigned to me	Fischer 2012
S10	I can achieve the objectives set for me	Fischer 2012
	Employee empowerment	
S11	I can decide on my own how to go about doing my work (e.g. order of tasks)	Spreitzer 1995
S12	I can make improvements to my way of working without checking first with my supervisor	Ryan et al. 1996
S13	I have a chance to participate in the decisions that affect my work	Schneider et al. 2003
S14	<Supplier name> is interested in my opinion concerning the cleaning service of Metsä Group's facilities (e.g. quality of work and service delivery)	Self-developed
	Employee engagement	
S15	I would recommend this company as a good organization to work for	Gelade & Young 2005
S16	I am proud to tell others that I work for <supplier name>	Gelade & Young 2005
S17	I am willing to work harder than is required of me	Sergeant & Frenkel 2000
S18	My job makes good use of my skills	Tornow & Wiley 1991
S19	I feel that <supplier name> appreciates my job	Hallowell et al. 1996
	Feedback	
S20	<Supplier name> asks our external customers to evaluate the quality of our cleaning service	Schneider et al. 2003
S21	We are informed about the results of external customer evaluations of the quality of our cleaning service	Schneider et al. 2003
S22	<Supplier name> adequately tracks the quality of its cleaning service	Schneider et al. 2003

S23	I receive regular feedback on my performance	Gelade & Ivery 2003
S24	I receive recognition when I do a good job	Hallowell et al. 1996
S25	I can improve the quality of my work based on the feedback I get	Self-developed
	Supplier's ability to develop the service	
S26	<Supplier name> is active in giving suggestions to its customers to improve the quality of cleaning service	Korpela 2015
S27	I feel encouraged to come up with better ways to do my job	Harmon et al. 2003
S28	In my working community suggestions for improvements are implemented	Fischer 2012
S29	At <supplier name> we obtain ideas from customers	Fischer 2012
S30	Ideas obtained from customers lead to improvements in our work	Self-developed

APPENDIX F: SURVEY ITEMS FOR THE BUYER COMPANY EMPLOYEES (CASE COMPANY SURVEY)

Coding	Factor and items	Original source
	Customer-employee interaction	
C1	The supplier employees are always willing to help me	Brady & Cronin 2001
C2	The supplier employees make the effort to understand my needs	Cronin et al. 2000
C3	The supplier employees seek the best for the customer	Caro & García 2008
C4	The supplier employees are friendly	Homburg & Garbe 1999
C5	The supplier employees are interested in our working community's opinion about cleaning service	Homburg & Garbe 1999
C6	The supplier employees inform our working community about problems concerning the cleaning service	Homburg & Garbe 1999
	Expertise	
C7	The supplier employees are competent	Cronin et al. 2000
C8	The supplier employees take initiative	Supplier customer satisfaction survey
C9	The behavior of the supplier employees is good	Hellier et al. 2003
C10	The appearance of the supplier employees is neat	Supplier customer satisfaction survey
	Responsiveness	
C11	The supplier employees perform the cleaning service promptly	Brady & Cronin 2001
C12	The supplier employees react to occurring problems	Ko & Pastore 2005
C13	The supplier employees react to Metsä Group's requests	Brady & Cronin 2001
C14	The supplier provides the cleaning service at the promised time	Hellier et al. 2003
	Perceived outcome quality	
C15	Generally, the quality of cleaning is as good as I expect	Homburg & Garbe 1999
C16	Overall, I'm satisfied with the cleanliness of the working spaces	Self-developed
C17	I'm satisfied with the cleaning service of the supplier	Ko & Pastore 2005
C18	The quality of the cleaning service of the supplier is so good, that I don't expect to find the same from other organizations	Salanova et al. 2005

APPENDIX G: SURVEY ITEMS FOR THE SUPPLIER-CUSTOMER RELATIONSHIP DIMENSION (SUPPLIER-CUSTOMER RELATIONSHIP SURVEY)

Coding	Factor and items	Original source
	Communication	
R1	We have meetings frequently enough with the <partner name>	Korpela 2015
R2	The aims of the meetings are reached	Korpela 2015
R3	The meetings with the <partner name> are useful	Korpela 2015
R4	We can solve problems concerning cleaning service together with the <partner name>	Korpela 2015
R5	Contacting the <partner name> representative is easy	Self-developd
R6	The <partner name> informs us about problems concerning the cleaning service	Homburg & Garbe 1999
	Trust	
R7	There is a mutual trust in the relationship	Korpela 2015
R8	The supplier keeps its promises to our company	Doney & Cannon 1997
R9	In our relationship with the supplier important decisions are taken jointly	Korpela 2015
R10	All issues are discussed openly in the relationship	Korpela 2015
R11	All issues are discussed confidentially in the relationship	Korpela 2015
R12	Both parties are committed to relationship development	Korpela 2015

APPENDIX H: REVIEW OF SERVICE QUALITY MODELS AND THE RESPECTIVE QUALITY DIMENSIONS

Study	**Quality dimensions**
Grönroos 1984	Technical quality Functional quality Image
Haywood-Farmer 1988	Physical process People's behavior Professional judgement
Parasuraman et al. 1988	Tangibles Reliability Responsiveness Assurance Empathy
Lehtinen & Lehtinen 1991	Three-dimensional: Physical quality Interactive quality Corporate quality OR Two-dimensional: Process quality Output quality
Cronin & Taylor 1992	Tangibles Reliability Responsiveness Assurance Empathy
Holmlund & Kock 1995	Economic Quality Functional Quality Technical Quality
Dabholkar et al. 1996	Physical aspects Reliability Personal interaction Problem solving Policy
Dabholkar et al. 2000	Reliability Personal attention Comfort Features
Brady & Cronin 2001	Interaction quality Service environment quality Outcome quality
Homburg & Garbe 2001	Structural quality Process-related quality Outcome-related quality
Gounaris 2005	Potential quality Hard process quality Soft process quality Output quality
Ko & Pastore 2005	Program quality Interaction quality Outcome quality Physical environment quality
Collier & Bienstock 2006	Process quality Outcome quality Recovery

Seth et al. 2006	Service reliability Credibility Service competence Intra-organisational communication Service flexibility Financial trust Pleasant enviroment
Caro & Garcia 2007	Personal interaction Design Physical environment Outcome

APPENDIX I: STUDIES OF SERVICE PROFIT CHAIN AND THE USED MEASURES FOR SUPPLIER PRACTICES

Source	Measures used for the supplier practices
Schneider et al. (1980)	Job satisfaction Organizational satisfaction Enthusiast orientation Bureaucrat orientation Effort rewarded Retain customers Personnel support Processing support Marketing support Equipment/supply support
Hallowell et al. (1996)	Internal service quality Tools Policies and procedures Teamwork Management support Goal alignment Effective training Communication Rewards and recognition Job satisfaction
Schneider et al. (1998)	Work facilitation Interdepartment service Global service climate Customer orientation Managerial practices Customer feedback
Schneider et al. (2003)	Satisfaction with empowerment Satisfaction with job fulfillment Satisfaction with pay Satisfaction with work group Satisfaction with security Satisfaction with work facilitation Overall job satisfaction
Gelade & Young (2005)	Climate Team climate Job enablers Support climate Commitment
Salanova et al. (2005)	Organizational resources Training Job autonomy Technology Work engagement Vigor Dedication Absorption Service climate
Schneider et al. (2005)	Service leadership behavior Service climate Customer-focused organizational citizenship behavior
Homburg et al. (2009)	Employee job satisfaction Employee-company identification